UNIVERSIDADE DE LISBOA

Faculdade de Medicina de Lisboa

USE OF GENOMIC DNA-REPORTER TOOLS TO DISSECT PATHOLOGICAL MECHANISMS CAUSED BY GAA EXPANSIONS IN FRIEDREICH'S ATAXIA

Ana Maria Ferreira da Silva

Orientadores: Professor Doutor Richard Wade-Martins

Professor Doutor Tiago Fleming Outeiro

Programa de Doutoramento em Neurociências

Tese especialmente elaborada para a obtenção do grau de Doutor em Ciências Biomédicas, especialidade de Neurociências

2015

UNIVERSIDADE DE LISBOA

Faculdade de Medicina de Lisboa

USE OF GENOMIC DNA-REPORTER TOOLS TO DISSECT PATHOLOGICAL MECHANISMS CAUSED BY GAA EXPANSIONS IN FRIEDREICH'S ATAXIA

Ana Maria Ferreira da Silva

Orientadores: Professor Doutor Richard Wade-Martins

Professor Doutor Tiago Fleming Outeiro

Programa de Doutoramento em Neurociências

Tese especialmente elaborada para a obtenção do grau de Doutor em Ciências Biomédicas, especialidade de Neurociências

Júri:

Presidente: Doutor José Luís Bliebernicht Ducla Soares, Professor Catedrático da Faculdade de Medicina da Universidade de Lisboa

Vogais: Professor Richard Wade-Martins, *Professor of Molecular Neuroscience* da Universidade de Oxford, United Kingdom; (*Orientador*)

Doutora Patrícia Espinheira Sá Maciel, Professora Associada da Escola de Ciências da Saúde da Universidade do Minho;

Doutor Luís Fernando Morgado Pereira de Almeida, Professor Auxiliar da Faculdade de Farmácia da Universidade de Coimbra;

Doutor Manuel Diamantino Pires Bicho, Professor Catedrático da Faculdade de Medicina da Universidade de Lisboa;

Doutor Mamede Alves de Carvalho, Professor Catedrático da Faculdade de Medicina da Universidade de Lisboa;

Doutor Tiago Fleming de Oliveira Outeiro, Professor Associado Convidado da Faculdade de Medicina da Universidade de Lisboa (*Co-orientador*).

Trabalho financiado pela Fundação para a Ciência e a Tecnologia através da bolsa de doutoramento SFRH/BD/61048/2009.

2015

As opiniões expressas nesta publicação

são da exclusiva responsabilidade do seu autor.

A impressão desta tese foi aprovada pelo Conselho Científico da Faculdade de Medicina de Lisboa em reunião de 15 de Dezembro de 2015.

To the loving memory of my grandfather, Francisco.

Acknowledgements

The work described in this Thesis would not have been possible without the support of my supervisors, colleagues, friends and family.

First and foremost, I would like to thank my supervisor Richard Wade-Martins for his time, encouragement, sharp insight and expectance of excellence. I am grateful that he entrusted me with the coolest project I could have ever hoped for, but also granted me enormous freedom to wander intellectually and pursue my own ideas. Without his support I would not have been able to see it through.

I would like to thank my co-supervisor Tiago Fleming Outeiro for his enthusiasm, constant encouragement and academic support throughout my PhD.

I am grateful for the outstanding support, knowledge and guidance of the talented Michele Lufino. I deeply enjoyed our intense and vibrant discussions which greatly boosted my intellectual growth.

My humble and sincere thanks to the exceptional minds of Jill Brown and Veronica Buckle for their scientific expertise and mentorship on the exciting field of nuclear organisation.

I am indebted to the past and present members of the Wade-Martins Laboratory for creating a supportive environment which allowed me to thrive. From a great team come greater friends: special thanks to Ruxandra Mutihac, Milena Cioroch, Heike Wobst, Tonya Taylor, Lawrence Tam and Lara Lourenço Venda.

I would like to acknowledge Eva Wegel and Ian Dobbie, from the Micron Facility, for providing microscopy training. I also thank Yaron Shav-Tal for the gift of NES-YFP-MS2-NLS construct, and Jonathan Chubb, Richard Parton and Timothy Weil for helpful discussions regarding the MS2 system.

I am grateful to Fundação para a Ciência e a Tecnologia for providing financial support during my PhD.

Lastly and most importantly, I thank my family for encouraging me in all of my pursuits and inspiring me to follow my dreams. In the words of Sir Isaac Newton, “If I have seen further, it is by standing on the shoulder of giants”.

Publications and conference presentations

Publications:

Silva AM, Brown JM, Buckle VJ, Wade-Martins R and Lufino MM (2015) Expanded GAA Repeats Impair *FXN* Gene Expression and Reposition the *FXN* Locus to the Nuclear Lamina in Single Cells. *Hum. Mol. Genet.* 24(12):3457-71.

Lufino MMP, **Silva AM,** Nemeth AH, Alegre-Albarrategui J, Russell AJ, Wade-Martins R (2013) A GAA Repeat Expansion Reporter Model of Friedreich's Ataxia Recapitulates the Genomic Context and Allows Rapid Screening of Therapeutic Compounds. *Hum. Mol. Genet.* 22(25): 5173-87.

Conferences:

Silva AM, Brown JM, Buckle VJ, Wade-Martins R and Lufino MM (2015) Expanded GAA Repeats Impair Frataxin Gene Expression and Promote Repositioning to the Nuclear Periphery at Single-Cell Level. *Oral presentation at the 2015 International Ataxia Research Conference, Windsor, UK.*

Silva AM, Groves J, Chubb JR, Outeiro TF, Wade-Martins R, Lufino M (2012) Live-cell visualization of the effect of a GAA repeat expansion on transcription of the *FXN* gene. *Poster presentation at the Ataxia Research Conference, Stanstead, UK.*

Silva AM, Lufino M, Groves J, Wade-Martins R (2012) Epigenetic silencing induced by a GAA repeat expansion in a cell model of Friedreich's ataxia. *Poster presentation at Epigenomics, Keystone, CO, USA.*

Silva AM, Lufino M, Russell A, Wade-Martins R (2011) Analysis of epigenetic mechanisms induced by GAA triplet expansion in a human cell model of Friedreich's ataxia. *Poster presentation at the FARA Research Conference, Strasbourg, France.*

Abstract

In Friedreich's ataxia (FRDA), abnormal GAA repeat expansions in intron 1 of the frataxin gene (*FXN*) cause epigenetic changes and reduce *FXN* mRNA levels in averaged cell samples though a poorly understood mechanism. Dissecting the silencing mechanism in FRDA *in situ* is crucial to improve our understanding of the disease.

Here, I use novel FRDA human cell models suitable for screening compounds able to upregulate *FXN* expression and to analyse the link between *FXN* nuclear localisation and expression in single cells. *FXN-Luc*, *FXN-GAA-Luc*, *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* stable human clones carry a site-specific integration of a single copy of the whole *FXN* locus with either 6 (*FXN-Luc* and *FXN-MS2-Luc*) or ~310 (*FXN-GAA-Luc* and *FXN-GAA-MS2-Luc*) GAA repeats in intron 1. To fluorescently label the transgenic *FXN* mRNA, I inserted MS2 binding sites into exon 2 of *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* transgenes by homologous recombination. The ~310 GAA repeats recapitulate the characteristic *FXN* gene repression and epigenetic changes seen in FRDA.

I report a single-cell analysis of *FXN* repression in which I identify the nuclear lamina (NL) as a novel and key player in *FXN* transcriptional impairment and silencing. Using a multidisciplinary approach, including analysis in both fixed and living single cells, I show that expanded GAA repeats increase *FXN* positioning at the NL, leading to decreased numbers of *FXN* mRNA molecules and slower transcription kinetics in the *FXN-GAA-MS2* cell model. Restoring histone acetylation reverses NL positioning. I observe the same abnormal repositioning to the NL in carrier and FRDA patient cells and show that this tightly correlates with a marked decrease in the number of actively expressing *FXN* alleles. Furthermore, I show that those few active expanded *FXN* alleles located at the NL express at a significantly lower level than the alleles located in the interior of the nucleus. Finally, I demonstrate that expanded GAA repeats predominantly disrupt *FXN* transcription initiation.

Collectively, these results suggest repressive epigenetic modifications at the expanded GAA-*FXN* locus may lead to NL relocation, where further repression may occur.

The mechanisms described may extend to other genetic diseases mediated by repeat expansions within regions of non-coding DNA.

Keywords: Friedreich's ataxia, GAA repeat expansion, single-cell resolution, heterochromatin, transcription.

Resumo

Na Ataxia de Friedreich (FRDA), uma expansão de repetições trinucleotídicas GAA, presentes no intrão 1 do gene que codifica a proteína frataxina (*FXN*), causa mudanças epigenéticas e reduz os níveis médios de RNA do gene *FXN* em amostras celulares através de um mecanismo pouco conhecido. Com o intuito de melhor perceber a patogénese inerente à doença e, em última análise, desenvolver terapias eficientes para a FRDA, é importante criar modelos celulares que traduzam as características repressivas da doença ao mesmo tempo que permitem quantificar eficientemente os níveis de expressão do gene *FXN*.

As linhas celulares reporter *FXN-GAA-Luc*, *FXN-Luc*, *FXN-GAA-MS2-Luc* e *FXN-MS2-Luc*, descritas nesta Tese, foram especificamente criadas de modo a permitir a comparação directa entre o efeito das repetições GAA normais e expandidas na expressão do gene *FXN*. Para o efeito usei: (i) todo o locus *FXN* com o seu promotor, intrões e exões originais e todos os elementos necessários para a expressão fisiológica do transgene, com inserção do gene *luciferase* no final do exão 5a; e (ii) uma única cópia de cada BAC integrado num sítio FRT especificamente localizado no cromossoma 1 de células HEK FRT, de modo a excluir efeitos contraditórios na expressão do gene *FXN* devido a integração de ambos os vectores em sítios diferentes.

Análise das modificações das histonas no promoter do gene *FXN* e regiões que ladeiam as repetições GAA a montante e a jusante revelou um decréscimo da acetilação de H3K9 e H4K8 e um aumento da metilação de H3K9me2 e H3K9me3 nas três regiões nas células *FXN-GAA-Luc*. Adicionalmente, as regiões a montante e a jusante das repetições GAA das células *FXN-GAA-Luc* apresentam um aumento da metilação do DNA em CpG específicos. As células *FXN-GAA-Luc* foram usadas num rastreio de compostos terapêuticos e permitiram identificar uma molécula capaz de aumentar a expressão do gene *FXN-GAA-Luc* para níveis similares aos das células *FXN-Luc*. A análise usando a técnica de imunoprecipitação da cromatina em células derivadas de pacientes depois de tratadas com esta molécula revelou o restauro para níveis normais de acetilação de H3K9 e H4K8 nas regiões que ladeiam as repetições GAA. Estes resultados sugerem que as repetições GAA induzem a repressão do gene *FXN* nas células *FXN-GAA-Luc* através da alteração da estrutura da cromatina no transgene, fazendo com que estas células sejam consideradas excelentes para o rastreio de moléculas capazes de aumentar a expressão do gene *FXN*.

No entanto, o modelo *FXN-GAA-Luc* apenas apresenta o estado provável do gene *FXN* visto que os resultados provêm de experiências onde se efectuam medições médias resultantes de amostras contento milhões de células. Consequentemente, o modelo *FXN-GAA-MS2-Luc* foi criado para dissecar o mecanismo repressivo de FRDA *in situ*, permitindo a visualisação e análise da localização e repressão de *FXN* em células fixas e vivas usando o sistema MS2.

Neste trabalho, desenvolvi um modelo celular humano para analisar a associação entre a localização e a expressão do gene *FXN* ao nível da célula. Os clones celulares estáveis *FXN-MS2-Luc* e *FXN-GAA-MS2-Luc* foram gerados por integração dos transgenes num local específico e contêm todo o locus *FXN* de 80 kb, o gene repórter luciferase no exão 5a e seis repetições ou uma expansão de ~310 tripletos GAA no intrão 1, respectivamente. Para efectuar uma marcação fluorescente do mRNA *FXN* transgénico e quantificar o efeito da expansão de tripletos GAA na transcrição do gene, inseri 24 repetições de locais de ligação da proteína MS2 (MBS) no exão 2 por recombinação homóloga. A expansão de ~310 GAA na linha celular *FXN-GAA-MS2-Luc* traduz as características repressivas do gene *FXN* em FRDA.

A localização do transgene *FXN* nas linhas *FXN-MS2-Luc* e *FXN-GAA-MS2-Luc* foi determinado por Immuno-FISH. *FXN* contactou com a lâmina nuclear (NL) em ~44% das células *FXN-GAA-MS2-Luc* quando comparado com apenas ~10% das células *FXN-MS2-Luc*. Após tratamento das células com inibidores das desacetilases de histonas, apenas o transgene *FXN-GAA-MS2-Luc* se reposicionou longe da NL. No entanto, ocorreu um aumento da expressão do mRNA *FXN* transgénico nas duas linhas celulares, sugerindo que existe uma associação complexa entre repressão do gene *FXN* e a sua localização intranuclear. Para aprofundar o conhecimento sobre a repressão do gene *FXN*, analisei o output de transcrição de alelos *FXN* transgénicos individuais nas células *FXN-MS2-Luc* e *FXN-GAA-MS2-Luc* por RNA FISH e recuperação de fluorescência após a fotodegradação (FRAP). As células *FXN-GAA-MS2-Luc* contêm ~5 ± 2 mRNAs por célula e as células *FXN-MS2-Luc* cells contêm ~9 ± 4 mRNA por célula, indicando que ~310 tripletos GAA reduzem o número de moléculas de mRNA em 44% ao nível celular. As curvas FRAP mostram que o tempo necessário para a recuperação completa da fluorescência após a fotodegradação é diferente nas duas linhas celulares. O transgene *FXN-MS2-Luc* apresentou um tempo total de recuperação de 120 segundos, enquanto o transgene *FXN-GAA-MS2-Luc* apresentou uma cinética mais lenta com recuperação total de 260 segundos. Devido ao facto de a estabilidade do mRNA de *FXN* não apresentar diferenças entre as

células *FXN-GAA-MS2-Luc* e *FXN-GAA-MS2-Luc*, os resultados de RNA FISH e FRAP indicam que a expansão de GAA diminui a quantidade de moléculas de mRNA de *FXN-GAA-MS2-Luc* através do impedimento da iniciação e/ou elongação da transcrição por parte da polimerase de RNA II. De modo a elucidar a relação entre a localização e repressão do gene *FXN*, analisei a intensidade fluorescente de transgene activos em células *FXN-MS2-Luc* e *FXN-GAA-MS2-Luc* vivas. A intensidade fluorescente de locais de transcrição foi significativamente menor quando os transgenes estavam a expressar na periferia nuclear comparando com o nucleoplasma em células *FXN-MS2-Luc* e *FXN-GAA-MS2-Luc*. Estes dados indicam que os dois transgenes expressam quando localizados na periferia nuclear, embora o façam em quantidades mais pequenas. Quando comparados com transgenes *FXN-MS2-Luc* activos, a intensidade dos locais de transcrição dos transgenes *FXN-MS2-GAA-Luc* foi significativamente mais baixa apenas no interior do núcleo. No seu conjunto, estes resultados sugerem que a expansão de GAA aumenta a localização do transgene *FXN-GAA-MS2-Luc* na NL, onde os níveis de expressão são reduzidos quando comparados com o interior nuclear ou com os níveis de expressão do transgene *FXN-MS2-Luc.*

Em seguida, analisei a localização do gene no seu ambiente genómico natural em células derivadas de pessoas saudáveis, de portadores heterozigóticos de um alelo *FXN* mutante e de pacientes. Em células derivadas de portadores heterozigóticos, o alelo *FXN* contendo a expansão de GAA localiza-se preferencialmente mais próximo da periferia nuclear do que o alelo normal e contacta mais vezes com a NL. Quando se compara a localização do gene *FXN* em células saudáveis com células derivadas de pacientes, os resultados indicam que a expansão de GAA aumenta a probabilidade de um alelo se encontrar associado à NL e consequentemente a probabilidade de estar silenciado. Adicionalmente, quando os alelos *FXN* expandidos se encontravam a expressar, os níveis de expressão eram significativamente reduzidos quando se encontram no nucleoplasma, mas especialmente quando localizados na periferia nuclear. Estes resultados indicam uma relação directa entre o posicionamento do gene *FXN* na NL e a repressão da transcrição mediada pela expansão de GAA. A realização de uma quantificação ao nível cellular mostrou ainda que a expansão das repetições de GAA induz um défice na expressão maioritariamente ao nível da iniciação da transcrição do gene *FXN*, mas também induz um pequeno bloqueio na elongação da polimerase de RNA II.

No seu conjunto, estes resultados sugerem que as modificações epigenéticas repressivas no locus *FXN* expandido podem induzir a relocalização do gene para a NL,

onde uma repressão adicional pode ocorrer. O efeito combinado da presença da expansão de GAA e relocalização do gene para a NL resultam numa redução catastrófica dos níveis de transcrição, levando à redução dos níveis da proteína frataxina e, em última análise, à manisfestação de FRDA. O trabalho descrito nesta Tese apresenta novos conhecimentos sobre as causas moleculares subjacentes à FRDA e poderá ser aplicável a outras doenças genéticas causadas por expansões de nucleótidos em regiões não codificantes do DNA.

Palavras-chave: Ataxia de Friedreich, expansão de repetições GAA, resolução ao nível da célula, heterocromatina, transcrição.

Contents

List of figures

List of Tables

List of movies (available on CD)

Movie 1: Live-cell imaging of pBAC-*FXN-MS2-Luc* transcription in HEK FRT cells.

Movie 2: Live-cell imaging of pBAC-*FXN-GAA-MS2-Luc* transcription in HEK FRT cells.

Movie 3: Real time visualization of *FXN-MS2-Luc* transcription.

Movie 4: Real time visualization of *FXN-GAA-MS2-Luc* transcription.

Movie 5: FRAP of a *FXN-MS2-Luc* transcription site.

Movie 6: FRAP of a *FXN-GAA-MS2-Luc* transcription site.

List of abbreviations

5hmC	5-hydroximethylcytosine
5mC	5-methylcytosine
AAV	Adeno-associated virus
ActD	Actinomycin D
AID	Activation-induced cytidine deaminase
ANOVA	Analysis of variance
ATP	Adenosine 5′-triphosphate
ATXN8	Ataxin 8 gene
ATXN8OS	Ataxin 8 opposite strand gene
BAC	Bacterial artificial chromosome
BAF	Barrier to autointegration factor
BCA	Bicinchoninic acid solution
Bp	Base pairs
BSA	Bovine serum albumin
BTF	BCL-2 associated transcription factor
CCD	Charge-coupled device
CFU	Colony forming units
ChIP	Chromatin immunoprecipitation
CMV	Human cytomegalovirus
CT	Chromosome territory
CTCF	CCCTC-binding factor
DAM	DNA adenine methyltransferase
DAPI	4′,6-diamidino-2- phenylindole
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
DM	Myotonic dystrophy
DMEM	Dulbecco’s modified Eagle’s medium
DMSO	Dimethyl sulphoxide
DNMT	DNA methyltransferase
dNTP	Deoxynucleotide triphosphate

DRG	Dorsal root ganglia
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EGR3	Early growth response factor 3
EGTA	Ethylene glycol tetraacetic acid
EMCCD	Electron multiplying charge-coupled device
FAST1	Frataxin antisense transcript 1
FBS	Fetal Bovine Serum
FISH	Fluorescence *in situ* hybridisation
FPN	Ferroportin
FRAP	Fluorescence recovery after photobleaching
FRAXA	Fragile X syndrome
FRAXE	Fragile X mental retardation associated with the *FRAXE* site
FRDA	Friedreich's ataxia
FRT	Flp recombinase target
FRTs	Ferritins
FTD/ALS	Frontotemporal dementia and amyotrophic lateral sclerosis
FXTAS	Fragile X tremor and ataxia syndrome
FXN	Frataxin gene
GCL	Germ-cell-less
GFAP	Glial fibrillary acidic protein gene
HAT	Histone acetyltransferase
HD	Huntington's disease
HDAC	Histone deacetylase
HDM	Histone demethylase
HEK	Human embryonic kidney 293 cells
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMT	Histone methyltransferase
HP1	Heterochromatin protein 1
HSV-1	Herpes simplex virus type 1
i-FXN	Intermediate frataxin
INM	Inner nuclear membrane
IP	Immunoprecipitation

iPSC	Induced pluripotent stem cell
ISCs	Iron-sulphur clusters
ISH	*In situ* hybridisation
K	Lysine
KDa	Kilodalton
KIKI	Knock-in mice
KIKO	Knock-in – knockout mice
LAD	Lamina-associated domains
LB	Luria broth
LBR	Lamin B receptor
LEM	Lap2β, Emerin and MAN1
LINEs	Long interspersed nuclear elements
LN	Long normal
MBS	MS2-binding site
MCP	MS2 coat protein
MCS	Multiple cloning site
MECP2	Methyl-CpG-binding protein 2
m-FXN	Mature frataxin
MIR	Mammalian-wide interspersed repeats
MMR	Mismatch repair proteins
MPP	Mitochondrial processing peptidase
MSH	MutS homologs
NA	Numerical aperture
NCoR	Nuclear corepressor complex
NES	Nuclear export signal
NL	Nuclear lamina
NLS	Nuclear localisation signal
NOR	Nucleolus organiser region
NP	Nuclear periphery
NTC	No template control
ONM	Outer nuclear membrane
ORF	Open reading frame
PCNA	Proliferating cell nuclear antigen

PBS	Phosphate-buffered saline
PcG	Polycomb group
PEV	Position effect variegation
PFA	Paraformaldehyde
PIP5K1B	phosphatidylinositol 4-phosphate 5-kinase β type I gene
PML	Promyelocytic leukemia
PCP	Phage PP7 coat protein
PCR	Polymerase Chain Reaction
PRC	Polycomb repressive complexes
R	Purine
RAN	Repeat-associated non-ATG
RISC	RNA-induced silencing complex
RNAi	RNA interference
RNAP	RNA polymerase
ROI	Region of interest
SAP	Shrimp alkaline phosphatase
SBMA	Spinal and bulbar muscular atrophy
SCA	Spinocerebellar ataxia
SDS	Sodium dodecyl sulfate
SEM	Standard error of the mean
SINEs	Short interspersed nuclear elements
SN	Short normal
snRNPs	small nuclear ribonucleoprotein particles
SRF	Serum responsive factor
SSC	Saline sodium citrate
TALE	Transcription activator-like effector protein
TBE	Tris-borate EDTA
TDG	Thymine DNA glycosylase
TE	Tris-EDTA
TET	Ten-eleven translocation methylcytosine dioxygenase
TFAP2	Transcription factor AP2
TFR1	Transferrin receptor 1
TSS	Transcription start site

UTR	Untranslated region
WT	Wild-type
Y	·Pyrimidine
YAC	Yeast artificial chromosome
Yfh1	Yeast frataxin homolog
ZFN	Zinc finger nucleases

CHAPTER 1

General introduction

1.1 Unstable expanded repeat diseases

Abnormally expanded DNA repeats are associated with several neurodegenerative diseases (Table 1.1). The majority of these diseases are caused by expanded trinucleotide repeats, but tetra-, penta-, and hexanucleotide repeat expansions have also been identified. Expanded repeats can be found either in coding or non-coding regions of a given locus. In the first group, exonic repeats code for amino acid homopolimers, as occurs in diseases mediated by polyglutamine and polyalanine runs in proteins, for example Huntington's disease (HD) and oculopharyngeal muscular dystrophy, respectively. In the second group, repeats can be located in: (i) the 5′ untranslated regions (5′ UTRs), as in fragile X syndrome (FRAXA), fragile X mental retardation associated with the *FRAXE* site (FRAXE), fragile X tremor and ataxia syndrome (FXTAS), and spinocerebellar ataxia (SCA) 12; (ii) the 3′UTRs, as in myotonic dystrophy (DM) 1, SCA8 and HD-like 2; (iii) introns, as in DM2, Friedreich's ataxia (FRDA), SCA10, SCA31; SCA36 and *C9ORF72*-related frontotemporal dementia and amyotrophic lateral sclerosis (FTD/ALS) (reviewed in Gatchel & Zoghbi, 2005; Mirkin, 2007; La Spada & Taylor, 2010; Polak *et al*, 2013).

Although each particular disease is caused by a mutation in a different locus, several features are shared. Alleles are present in three states: normal, premutated and fully mutated. While healthy individuals carry less than 30 repeats, patients harbor longer and unstable tracts of repeats on mutant alleles, with a strong tendency to expand rather than contract. Repeats show intergenerational and somatic instability. As longer repeat tracts are transmitted through generations, disease symptoms become more severe and appear at an earlier age, a phenomenon known as anticipation (reviewed in Dion & Wilson, 2009; Castel *et al*, 2010). Repeat instability seems to be triggered by formation of unusual DNA structures, such as hairpins, slipped-strand structures, triplexes and quadruplexes, during DNA replication, repair, recombination and transcription, where there is a transient separation of complementary DNA strands, or exposure of a single DNA strand. The DNA repair machinery is then thought to bind to these aberrant structures, stabilizing instead of repairing them, which results in repeat expansions or contractions (Mirkin, 2007; Polak *et al*, 2013). Changes in the epigenetic landscape of the mutated locus, such as aberrant DNA methylation, histone modifications, and chromatin remodelling also play a role in the molecular pathogenesis of several unstable expanded repeat diseases (reviewed in Evans-Galea *et al*, 2013).

Table 1.1 List of unstable repeat expansion neurodegenerative diseases (Gatchel & Zoghbi, 2005; La Spada & Taylor, 2010; Walsh *et al*, 2014)**.**

				Repeat length	
Disease	**Main clinical features**	**Mutation**	**Location**	**Normal**	**Expanded**
Loss of protein function					
FRDA	Ataxia, sensory loss, weakness, diabetes mellitus, cardiomyopathy	$(GAA)_n$	*FXN* (Intron 1)	6–32	70–1700
FRAXA	MR, facial dysmorphism, autism	$(CGG)_n$	*FMR1* (5′ UTR)	6–52	>200
FRAXE	MR, hyperactivity	$(GCC)_n$	*FMR2 (5′* UTR)	4–39	200–900
Toxic gain of RNA function					
SCA10	Ataxia and seizures	$(ATTCT)_n$	*ATXN10* (Intron 9)	10–29	400–4500
SCA31	Ataxia, dysarthria, nystagmus	$(TGGAA)_n$	*TK2/BEAN* (Intron)	0	≥100
SCA36	Ataxia, eye movement abnormalities, tongue fasciculations, upper and lower motor neuron disease	$(GGCCTG)_n$	*NOP56* (Intron 1)	3–14	650–2500
DM1	Muscle weakness, myotonia, cardiac conduction deficit, MR	$(CTG)_n$	*DMPK* (3′ UTR)	5–37	50–1000
DM2	Muscle weakness, myotonia, cardiac conduction deficit	$(CCTG)_n$	*ZNF9* (Intron 1)	10–26	75–11,000
FXTAS	Ataxia, intention tremor, parkinsonism, cognitive deficit	$(CGG)_n$	*FMR1* (5′ UTR)	6–52	55–200
Toxic gain of protein function					
DRPLA	Ataxia, epilepsy, choreoathetosis, dementia	$(CAG)_n$	*ATN1* (Exon 5)	7–34	49–88
HD	Severe movement abnormalities, chorea, dystonia, cognitive decline, psychiatric features	$(CAG)_n$	*HTT* (Exon 1)	6–28	36–180
SBMA	Motor weakness, swallowing difficulty, gynecomastia, hypogonadism	$(CAG)_n$	*AR* (Exon 1)	9–36	38–62
SCA1	Ataxia, dysarthria, spasticity, cognitive impairment	$(CAG)_n$	*ATXN1* (Exon 8)	6–39	41–83
SCA2	Ataxia, slow eye movement, hyporeflexia, polyneuropathy, motor neuropathy, infantile variant	$(CAG)_n$	*ATXN2* (Exon 1)	15–24	32–200
SCA3 (MJD)	Ataxia, dystonia, lower motor neuron disease	$(CAG)_n$	*ATXN3* (Exon 10)	13–36	61–84

SCA6	Ataxia, dysarthria, nystagmus, tremor	$(CAG)_n$	*CACNA1A* (Exon 47)	4–20	20–33
SCA7	Ataxia, retinal degeneration, cardiac involvement in infantile variant	$(CAG)_n$	*ATXN7* (Exon 3)	4–35	37–306
SCA17	Ataxia, behavioural changes or psychosis, intellectual deterioration, seizures	$(CAG)_n$	*TBP* (Exon 3)	25–44	47–63
Toxic gain of function at both the protein and RNA levels					
SCA8	Ataxia, dysarthria, nystagmus	$(CTG)_n$*$(CAG)_n$	*ATXN8OS* (3′ UTR) and *ATXN8* (Exon 1)	15–50	71–1300
Haploinsufficiency and toxic gain of function at both the protein and RNA levels					
FTD/ALS	Cognitive impairment, psychosis/Upper and lower motor neuron disease, paralysis	$(GGGGCC)_n$	*C9ORF72* (5'UTR/Intron 1)	2–19	30–2000
Unknown pathogenic mechanism					
SCA12	Ataxia, tremor, dementia	$(CAG)_n$	*PPP2R2B* (5′ UTR)	7–28	55–78
HDL2	Similar to HD	$(CTG)_n$*$(CAG)_n$	*JPH3* (exon 2A)	7–28	66–78

Abbreviations: *AR*, androgen receptor; *ATXN*, ataxin; *BEAN*, brain-expressed, associated with Nedd4; *CACNA1A*, calcium channel, voltage-dependent, P/Q type, α1A subunit; DM1, myotonic dystrophy type 1; DM2, myotonic dystrophy type 2; *DMPK*, dystrophia myotonica protein kinase; DRPLA, dentatorubral-pallidoluysian atrophy; DM, dystrophia myotonica; *FMR1*, fragile X mental retardation 1; *FMR2*, fragile X mental retardation 2; FRAXA, fragile X syndrome; FRAXE, fragile X mental retardation associated with the *FRAXE* site; FRDA, Friedreich ataxia; FTD/ALS, frontotemporal dementia and amyotrophic lateral sclerosis; *FXN*, frataxin; FXTAS, fragile X tremor/ataxia syndrome; *JPH3*, junctophilin 3; HD, Huntington's disease; HDL2, Huntington's disease-like 2; *HTT*, huntingtin; MJD, Machado–Joseph disease; MR, mental retardation; *NOP56*, nucleolar protein 56; *PPP2R2B*, protein phosphatase 2 (formerly 2A) regulatory subunit B; SBMA, spinal and bulbar muscular atrophy; SCA, spinocerebellar ataxia; *TBP*, TATA box binding protein; *TK2*, thymidine kinase 2; *ZNF9*, zinc-finger protein 9.

There are three key pathological mechanisms by which expansions result in disease: (i) loss of function typically by disrupting transcription of the mutated gene (FRDA, FRAXA, FRAXE); (ii) RNA toxic gain of function through aberrant interactions of the respective RNAs with proteins controlling RNA metabolism (SCA10, SCA31, SCA36, DM1, DM2, FXTAS); and (iii) protein toxic gain of function via abnormal protein interactions and accumulation of aggregates (polyglutamine diseases, for example HD) (reviewed in Nelson *et al*, 2013). Interestingly, some diseases have more than one repeat-mediated mechanism at play, for example SCA8 and *C9ORF72*-related FTD/ALS. In SCA8, two overlapping genes expressed in opposite directions are affected by the expansion, with a $(CAG)_n$ expansion in the ataxin 8 gene (*ATXN8*) and a $(CTG)_n$ expansion in the non-coding 3' UTR of the ataxin 8 opposite strand gene (*ATXN8OS*). As a result, *ATXN8* expresses a polyglutamine expansion protein, and *ATXN8OS* produces non-coding

CUG expansion mRNAs which accumulate as RNA foci and sequester RNA-binding proteins (Moseley *et al*, 2006; Daughters *et al*, 2009). In *C9ORF72*-related FTD/ALS, the $(GGGGCC)_n$ hexanucleotide repeat expansion leads to disease by: (i) formation of DNA and RNA G-quadruplexes and RNA/DNA hybrids (R-loops) which lead to abortive transcription; (ii) formation of RNA foci that bind to ribonucleoproteins, preventing RNA-processing factors from functioning normally in the nucleus; and (iii) repeat-associated non-ATG (RAN) translation, which results in the production of dipeptide repeat proteins that also form inclusions in affected cells (DeJesus-Hernandez *et al*, 2011; Renton *et al*, 2011; Mori *et al*, 2013; Haeusler *et al*, 2014).

In the following sections, I review the major findings in FRDA and also highlight the similarities with other repeat expansion diseases with special emphasis on non-coding repeats at the DNA and RNA level.

1.2 Friedreich's ataxia

Friedreich's ataxia (FRDA; OMIM 229300) is a progressive neurodegenerative disease and the most common form of recessive ataxia (Campuzano *et al*, 1996). The disease is named after Nikolaus Friedreich (1825–1882), a German pathologist and neurologist, who first described its clinical features (Friedreich, 1863a, 1863b, 1863c) and hereditary nature (Friedreich, 1876, 1877). An abnormal GAA trinucleotide–repeat expansion in intron 1 of the frataxin gene (*FXN*) causes FRDA. Approximately 96% of FRDA patients are homozygous for a GAA repeat expansion and the remaining patients (4%) are compound heterozygous with one expanded allele and a classical mutation (nonsense, missense, deletions, insertions) in the second allele (Campuzano *et al*, 1996; Bidichandani *et al*, 1997; Cossée *et al*, 1999). *FXN* alleles in healthy individuals contain <36 GAA repeats, whereas in FRDA patients GAA expansions ranging from ~70 to 1700 GAA repeats lead to *FXN* mRNA deficiency and subsequent reduced levels of frataxin, a nuclear-encoded mitochondrial protein essential for life (Campuzano *et al*, 1996; Cossée *et al*, 2000).

1.2.1 Etiology

Normal *FXN* alleles are divided in 2 groups: short normal (SN, <10 GAA repeats) and long normal (LN, from 12 to 60 GAA repeats). LN *FXN* alleles account for ~17% of normal alleles. Haplotype analysis of five markers (FAD1-ITR4-F5225-ITR3-CS2) close to the GAA repeats revealed that a single founder event that occurred on the major AT2CC

haplotype is at the origin of the LN alleles and the expanded alleles. The two second most frequent haplotypes AT3CC and AT2CT are most likely derived from the major haplotype by marker mutation rather than by independent events on different haplotypes (Cossée *et al*, 1997). Expanded *FXN* alleles are thought to have arisen through a two-step process. A singular event was responsible for the transition of SN to LN GAA repeats as these alleles only share the same haplotype rarely. This is thought to have occurred through DNA polymerase "stuttering" during DNA replication, originating LN alleles of $(GAA)_{18}$ by duplication of SN alleles of $(GAA)_8$. After a number of small increases in size because of slippage events, larger LN alleles served as a pool for further repeat expansions once the instability threshold (~34 GAA repeats) was reached (Montermini *et al*, 1997a; Labuda *et al*, 2000). Therefore, repeat numbers between 35 and 66 are referred as premutations. In line with this, all LN and expanded alleles share a common haplotype (Cossée *et al*, 1997).

A single initial event is also thought to have occurred in DM1, HD, DRPLA and MJD where haplotypes associated preferentially or uniquely with expanded alleles are also preferentially associated to LN alleles (Imbert *et al*, 1993; Rubinsztein *et al*, 1995; Takiyama *et al*, 1995; Yamagata *et al*, 1996; Yanagisawa *et al*, 1996). Furthermore, as observed in FRDA, the prevalence of DM1, HD and DRPLA in a given population is correlated with the frequency of the LN alleles (Novelli *et al*, 1994).

The FRDA founding mutation is dated between 9,000 and 24,000 years ago and went through a population bottleneck (Colombo & Carobene, 2000) which coincides with the glacial period in Europe. This may provide an explanation for the highest prevalence of FRDA in the surviving northern Spain ice age refuge (Polo *et al*, 1991). Further support comes from the observed distribution of FRDA in the remaining European regions where the prevalence decreases from west to east and correlates with the frequency gradient of the chromosomal R1b marker, a genetic marker for the population that originated from the Cantabrian refuge (Tambets *et al*, 2004). It has been hypothesized that the FRDA distribution in Europe is derived from Palaeolithic migrations out of the Franco-Cantabrian ice age refuge (Vankan, 2013).

1.2.2 Epidemiology

FRDA is detected in individuals from Europe, the Middle East, North Africa and India (Labuda *et al*, 2000). The prevalence of the disease is closely correlated with the frequency of LN *FXN* alleles in the population, and is highest in Western Europe, with more than 1 case per 30,000 individuals (1/30,000) reported in north-west Spain and

Ireland (Schulz *et al*, 2009; Vankan, 2013). LN alleles are virtually absent in individuals from East Asia and in American Indians (Labuda *et al*, 2000).

The frequency of heterozygous mutation carriers in Europe shows a decreasing gradient from south to north-east, ranging from 1:78 in France to 1:89 in Germany and 1:500 in Finland (Cossée *et al*, 1997; Epplen *et al*, 1997; Labuda *et al*, 2000; Juvonen *et al*, 2002; Schulz *et al*, 2009; Vankan, 2013). The prevalence of FRDA in Caucasians is shown in Table 1.2.

Table 1.2 Prevalence of FRDA in Caucasians (source: Schulz et al, 2009; Vankan, 2013).

Country	Prevalence of FRDA
Cantabria, Spain	4.7/100,000
Denmark	0.5-1/100,000
Finland	0.13/100,000
France	4.2/100,000
Germany	3.1/100,000
Norway	1/100,000
Portugal	0.95/100,000
Sweden	0.23/100,000
United Kingdom	1.8/100,000

1.2.3 Clinical features of FRDA

The healthy *FXN* gene is expressed in all cells but at variable levels. In adult humans, *FXN* mRNA is most abundant in the heart and spinal cord, mild in the cerebellum, pancreas, liver, brown fat and skeletal muscle, and low in other tissues, such as the cortex (Campuzano *et al*, 1996; Koutnikova *et al*, 1997). In FRDA, residual levels of frataxin vary according to the length of the expansion and the cell type. Affected cells, such as sensory neurons and cardiomyocytes, are known to have a high mitochondrial content and high metabolic rate, making them more sensitive to frataxin deficiency.

The first neurologic symptoms usually appear around puberty, but early-onset and late-onset variants also exist. Patients experience progressive gait and limb ataxia, dysarthria and lower limb areflexia, resulting from the loss of large sensory neurons in the dorsal root ganglia (DRG), atrophy of the large sensory fibres in peripheral nerves, degeneration of the posterior columns, spinocerebellar and corticospinal motor tracts in the

spinal cord and dentate nucleus in the cerebellum (Fig. 1.1). Hearing disability and loss of vision are also common. Non-neurological symptoms include hypertrophic cardiomyopathy, diabetes mellitus, kyphoscoliosis and pes cavus. Life expectancy is reduced to an average of 40-50 years (Filla *et al*, 1990; Dürr *et al*, 1996; Pandolfo, 2009; Schulz *et al*, 2009).

In most cases, compound heterozygous patients are clinically indistinguishable from patients that are homozygous for the GAA repeat expansions, but a few missense mutations (e.g., G130V, D122Y, R165P, L106S) may cause an atypical or milder clinical phenotype due to an adverse effect on protein expression, function and/or stability (Cossée *et al*, 1999; Gellera *et al*, 2007).

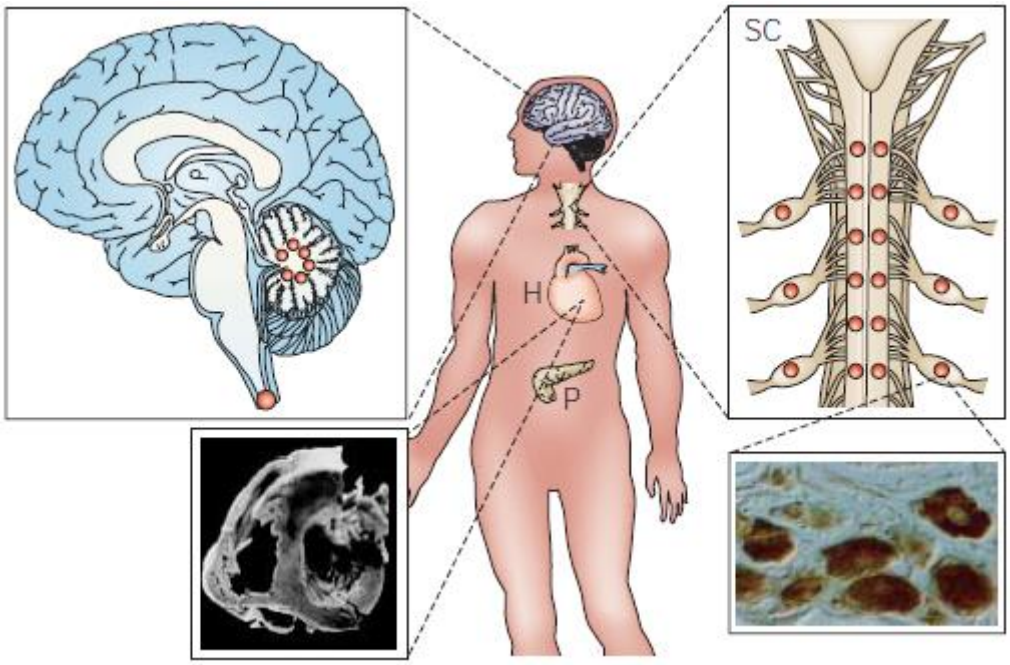

Figure 1.1 Main sites of organ dysfunction in FRDA. Large dots indicate more severe neuronal loss. Pathological features are shown for the brain, spinal cord (SC), heart (H) and pancreas (P). Loss of large primary neurons in dorsal root ganglia is a prominent and early hallmark. As the disease progresses, the posterior columns of the spinal cord degenerate. This associated with atrophy of the spinocerebellar tracts, the corticospinal motor tracts of the spinal cord and the large sensory fibers in peripheral nerves. Non-neurological symptoms include hypertrophic cardiomyopathy and diabetes (adapted from Taroni & DiDonato, 2004).

In homozygous patients, the GAA repeat length of the shorter of the 2 *FXN* expanded alleles correlates with residual frataxin protein levels, earlier onset (Fig. 1.2) and confinement to a wheelchair, and increased phenotypic severity of the disease (Dürr *et al*, 1996; Filla *et al*, 1996; Montermini *et al*, 1997b). The GAA expansion length only

accounts for ~50% of the variability in age of onset (Filla *et al*, 1996). Phenotypic variability among FRDA patients and within affected families can be also due to intergenerational instability of the GAA expansion, somatic mosaicism, interruptions in the repeat sequence, changes in expansion size over life and other modifying genes or environmental factors (Dürr *et al*, 1996; Montermini *et al*, 1997b; Pandolfo, 2009).

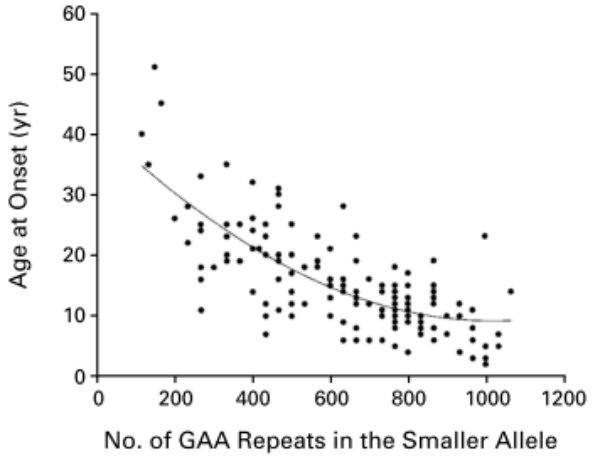

Figure 1.2 Correlation between age at onset and the number of GAA repeats in the smaller *FXN* allele. Larger GAA expansions in the smaller *FXN* allele correlates with earlier age at onset (Dürr *et al*, 1996).

For most of the unstable expanded repeat diseases, the repeat sequence shows intergenerational and age-dependent somatic instability. In FRDA, DM1, FRAXA and SCA8, expansions and contractions are equally likely after maternal transmission, whereas contractions are favoured after paternal transmission (Malter *et al*, 1997; Monrós *et al*, 1997; Pianese *et al*, 1997). In FRDA, expanded alleles are highly unstable in somatic cells *in vivo*. While expanded *FXN* allele with <250 GAA repeats tend to expand, *FXN* alleles with >500 GAA repeats show contractions in peripheral leukocytes from FRDA patients. Due to somatic instability, carriers carrying LN *FXN* alleles (44–66 triplets) can develop a FRDA phenotype (Sharma *et al*, 2002, 2004). Somatic instability in FRDA is mostly postnatal, progresses throughout life and is tissue-specific (De Biase *et al*, 2007a, 2007b). Expanded GAA repeats undergo progressive expansion in tissues primarily affected in FRDA: the DRGs, cerebellum and heart of FRDA patients (De Biase *et al*, 2007a), and in the DRGs, cerebellum and brain of mouse models (Al-Mahdawi *et al*, 2004; Clark *et al*, 2007; Anjomani Virmouni *et al*, 2015). Therefore, the progressive accumulation of large expansions coupled with a lower frequency of large contractions in specific cell types, such as DRGs, suggests that somatic instability may contribute to the selective vulnerability of

these cells in FRDA. Although the molecular mechanism of somatic instability in FRDA is still poorly understood, a few studies have suggested that mismatch repair (MMR) proteins may be involved (section 1.3.2.1.1).

In FRAXA and DM1, large expansions arise in non-dividing oocytes during arrest in meiotic prophase I, and in terminally differentiated somatic cells in females; in males, the full mutation alleles tend to contract in the dividing male spermatogonia, during weeks 13–17 of fetal development, and small gains and losses in the pre-mutation range can be observed in both spermatogonia and in somatic cells (McMurray, 2010).

1.3 The *FXN* locus

1.3.1 *FXN* in healthy individuals

The human *FXN* locus is located on the positive strand of chromosome 9q21.11 and contains seven exons (1–4, 5a, 5b, and non-coding 6). Alternative splicing results in multiple transcript variants with exons 1–5a generating the 1.3 kb major transcript that encodes the 210-amino acid frataxin protein (Campuzano *et al*, 1996). Two major transcription start sites (TSS) were identified: TSS1 is located 221 bp upstream of the ATG translation start site (Campuzano *et al*, 1996), whereas TSS2 is located 62 bp upstream of the ATG and is considered the major TSS in Epstein Barr virus-transformed lymphoblastoid cell lines. (Kumari *et al*, 2011).

The *FXN* promoter (Fig. 1.3) is enriched in repetitive DNA elements. These include L2 (LINE) and Alu (SINE) elements as well as MIRb and mariner DNA transposon downstream of TSS1 and in intron 1 (Greene *et al*, 2005). The presence of these different retroelements suggest that the *FXN* promoter sequence has been remodelled repeatedly during the evolution of mammals and fairly recently in the primate lineage (Greene *et al*, 2005). Although their contribution to *FXN* regulation is unclear, deleting these elements significantly impairs *FXN* expression (Greene *et al*, 2005). Alu and MIR elements are transcribed from internal RNA polymerase (RNAP) III promoters and may exert a positive *cis*-acting enhancer effect on RNAPII-transcribed genes (Oliviero & Monaci, 1988; Tomilin, 1999). Moreover, Alu elements are also thought to affect gene expression by altering nucleosome positioning (Englander & Howard, 1995) and providing direct binding sites for hormone receptors (Norris *et al*, 1995) and transcription factors (Deininger, 2011). The region between 121-221 bp upstream of the start of the frataxin open reading frame (ORF) shows homology between primates and rodents and contains important regulatory

elements for *FXN* expression (Greene *et al*, 2005). Interestingly, the *FXN* promoter does not present typical motifs of most mammalian promoters, such as a TATA box, and is not controlled by either the initiator or the downstream promoter-like elements found 24 bp downstream of TSS1 (Greene *et al*, 2005; Kumari *et al*, 2011).

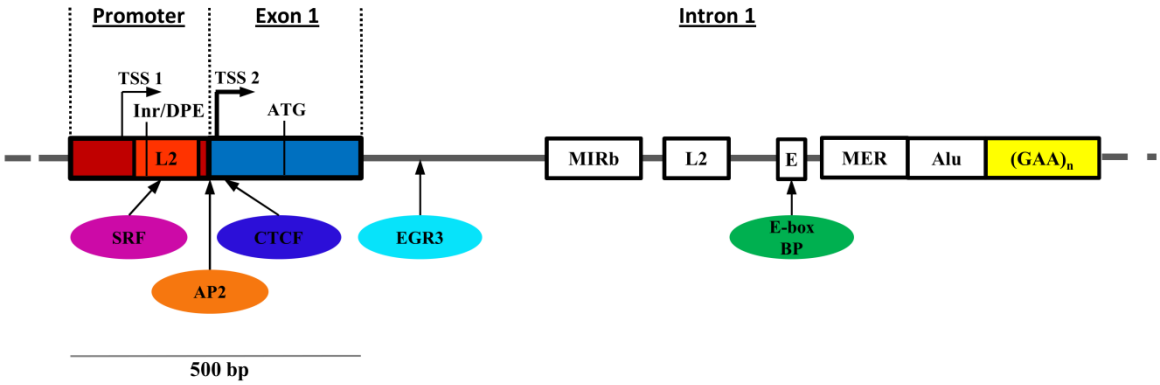

Figure 1.3 The 5' end of the *FXN* gene. Schematic representation of the the 5'end *FXN* gene showing its regulatory elements and protein binding sites. The repetitive DNA elements identified at the *FXN* locus include L2 (LINE) and Alu (SINE) elements as well as MIRb and mariner DNA transposon (MER). *FXN* is transcribed from TSS1, which is located 221 bp upstream of the ATG, and TSS2, located 62 bp upstream of the ATG. TSS2 is the major TSS in lymphoblastoid cell lines. The region between TSS1 and exon 1 is thought to be a TATA-less downstream promoter, which contains the initiator/downstream promoter-like elements (Inr/DPE). Binding sequences were identified for transcription factors SRF, TFAP2 and EGR3, and for the insulator protein CTCF. An E-box is present in the region upstream of the GAA repeats and can potentially be a target for MyoD or c-myc (adapted from Kumari & Usdin, 2012).

A few studies have identified transcription factors that influence *FXN* expression (reviewed in Yandim *et al*, 2013). Li and colleagues showed that SRF (serum responsive factor) and TFAP2 (transcription factor AP2) directly bind to sequences within and immediately downstream of the *FXN* promoter region conserved between rodents and primates due to the presence of the L2 retrotransposon-like sequence (Greene *et al*, 2005; Li *et al*, 2010). TFAP2 upregulates *FXN* mRNA expression in several cell lines (Li *et al*, 2010) and is important in the balance between cell proliferation and differentiation during embryogenesis (Eckert *et al*, 2005). SRF increases *FXN* expression only in specific cell-lines and activate genes involved in synaptic activity and plasticity (Knöll & Nordheim, 2009). Moreover, the chromatin insulator CCCTC-binding factor (CTCF) binds between 154–173 bp downstream of TSS1 (De Biase *et al*, 2009). CTCF is implicated in diverse roles in gene regulation, including context-dependent promoter activation/repression,

enhancer blocking and/or barrier insulation, hormone-responsive silencing, genomic imprinting, and long-range chromatin interactions (reviewed in Phillips & Corces, 2009).

The intronic region downstream of exon 1 is also important for *FXN* regulation. Deletion of the binding site for the early growth response factor EGR3 (Li *et al*, 2010) or of an E-box element (Greene *et al*, 2007) induced a decrease in *FXN* expression using reporter assays. The E-box element can potentially be bound by proteins of the large family of basic Helix-Loop-Helix that includes the muscle-specific factor MyoD and c-myc (Greene *et al*, 2007).

A polymorphic $(GAA)_n$ repeat sequence located approximately 1.3 kb downstream of TSS2 is embedded in the centre of an Alu element in intron 1 and is preceded by a poly(A) tract characteristic of Alu elements. *FXN* alleles in healthy individuals contain <36 GAA repeats, whereas in FRDA patients GAA expansions ranging from70 to 1700 GAA repeats lead to *FXN* mRNA deficiency (Campuzano *et al*, 1996; Cossée *et al*, 1997; Clark *et al*, 2004; Monticelli *et al*, 2004).

1.3.2 Molecular basis of FRDA

1.3.2.1 GAA expansion-mediated transcriptional dysregulation

In FRDA, expanded GAA repeats in intron 1 of the *FXN* gene reduce steady-state *FXN* mRNA levels to 4% to 29% of normal (Campuzano *et al*, 1997) through a still poorly understood mechanism. This severe decrease in *FXN* mature mRNA content could potentially be due to impairment in transcription, pre-mRNA processing or decay.

RNA splicing differences between normal and expanded *FXN* alleles have only been shown when using a *FXN* minigene construct containing a human cytomegalovirus (CMV) promoter, *FXN* exon 1, part of intron 1 and all of exon 2, transfected into mammalian cells (Baralle *et al*, 2008); or when inserting expanded GAA repeats into an intron of the *URA3* gene in yeasts (Shishkin *et al*, 2009). However, these results failed to replicate when studying the endogenous *FXN* transcript in FRDA patient cells (Bidichandani *et al*, 1998; Punga & Bühler, 2010). Moreover, there is no evidence showing differences in the decay rate of mature *FXN* transcripts from normal and expanded *FXN* alleles (Punga & Bühler, 2010).

Therefore, it seems plausible to assume that the *FXN* mRNA decrease seen in FRDA results from *FXN* transcriptional impairment. However, how the mutation interferes with the transcription machinery at the *FXN* gene remains unclear. Two non-exclusive

hypotheses have been proposed (reviewed in Kumari & Usdin, 2012; Yandim *et al*, 2013). Firstly, expanded GAA repeats may form non-B DNA structures, such as triplexes or sticky DNA (Bidichandani *et al*, 1998; Sakamoto *et al*, 1999), and RNA/DNA hybrids (R-loops; Grabczyk *et al*, 2007; Groh *et al*, 2014) which impair RNAPII elongation. Secondly, GAA repeat expansions can induce the formation of heterochromatin (Saveliev *et al*, 2003; Herman *et al*, 2006), leading to increased DNA methylation at specific CpG sites (Greene *et al*, 2007; Al-Mahdawi *et al*, 2008; Evans-Galea *et al*, 2012), reduced histone acetylation (H3/H4ac) and increased levels of methylated histones H3K9me2 and H3K9me3 (Herman *et al*, 2006; Al-Mahdawi *et al*, 2008). It has been suggested that these epigenetic changes surrounding the GAA expansion impair RNAPII elongation (Punga & Bühler, 2010), but also spread upstream towards the *FXN* promoter, inducing a non-permissive chromatin configuration for transcription initiation, altering nucleosome positioning and preventing insulator protein CTCF binding (De Biase *et al*, 2009; Kumari *et al*, 2011; Chutake *et al*, 2014; Groh *et al*, 2014).

1.3.2.1.1 Abnormal DNA structures

Repetitive DNA sequences are prone to adopting unusual DNA secondary structures due to sequence symmetry, base composition, DNA supercoiling and cellular ambient conditions (Mirkin, 2007). These structures contribute to the instability of disease-associated repeats by impacting on DNA metabolism, for example stalling replication, altering repair or causing unusual recombination, inducing the formation of hairpins, triplexes and quadruplexes. During transcription, repeats can form RNA/DNA hybrids (Castel *et al*, 2010).

Of all the repeat expansions associated with unstable expanded diseases, only GAA repeats do not form hairpins. Single-stranded CNG repeats form hairpin structures that consist of both Watson–Crick base pairs and mismatched base pairs, with a decreasing stability in the order CGG > CTG > CAG = CCG. In addition to hairpins, single-stranded CGG, CCG and GGGGCC repeats can fold into G-quadruplexes (Fry & Loeb, 1994; Marquis Gacy *et al*, 1995; Haeusler *et al*, 2014).

Expanded GAA repeats form non-B DNA conformations, such as triplexes, that can affect *FXN* transcription in a length and orientation-dependent manner (Bidichandani *et al*, 1998; Ohshima *et al*, 1998). Triplexes are three-stranded DNA structures formed at polypurine·polypyrimidine (R·Y) tracts, where the third strand forms Hoogsteen base pairs between the R or Y bases with the purines already paired with pyrimidines in the duplex

DNA. R·R·Y triplexes, unlike Y·R·Y, form more rapidly and at neutral pH. In FRDA, GAA repeat expansions form a R·R·Y triplex containing two GAA repeat strands along with a single strand of TTC repeats (Mariappan *et al*, 1999; Grabczyk & Usdin, 2000; Potaman *et al*, 2004). These structures are stabilised by negative supercoiling and can self-associate, forming complexes known as sticky DNA which can only be resolved at high temperatures (80°C) and with the addition of EDTA to remove divalent metal ions (Sakamoto *et al*, 1999; Wells, 2008). Sticky DNA can impede virtually all biological processes *in vitro*, including transcription, replication, repair and recombination (Wells, 2008).

Another model suggests that expanded GAA repeats can form R-loops, a shared feature with CGG, CCG, CAG, CTG and GGGGCC repeats (Reddy *et al*, 2011; Colak *et al*, 2014; Groh *et al*, 2014; Reddy *et al*, 2014). These RNA/DNA hybrids are formed during transcription when the nascent RNA hybridizes to the DNA template behind the elongating RNAPII. Indeed, R-loops have been shown to form at expanded GAA repeats following *in vitro* transcription (Reddy *et al*, 2011), in *E. coli* (Grabczyk *et al*, 2007), and on endogenous *FXN* alleles from FRDA patients cells (Groh *et al*, 2014).

All these unusual structures could impair *FXN* transcription by creating a physical blockage to the RNAPII elongation through the GAA repeat sequence and act as an initial trigger to promote *FXN* silencing (Sakamoto *et al*, 2001; Groh *et al*, 2014).

Repeat-mediated aberrant DNA structures may be subsequently processed by proteins of the DNA repair machinery, potentially inducing repeat instability (reviewed in Castel *et al*, 2010; Iyer *et al*, 2014). In healthy conditions, these proteins stabilize the genome by correcting DNA replication errors, attenuating chromosomal rearrangements, and mediating the cellular response to certain types of DNA damage. Eukaryotic MutS homologs (MSHs) detect specific lesions by distinct heterodimeric complexes: human MSH2 forms a heterodimer in solution with MSH6 (MutSα) or MSH3 (MutSβ). MutSα recognizes and initiates removal of base–base mispairs, a subset of insertion–deletion loops and certain types of lesions caused by DNA damaging agents, whereas MutSβ almost exclusively recognizes loops. Mismatch recognition by MutSα or MutSβ is followed by recruitment and activation of a latent endonuclease function in MutLα (heterodimer of MLH1 and PMS2) in the presence of ATP and DNA-loaded replication sliding-clamp proliferating cell nuclear antigen (PCNA), resulting in single-strand breaks flanking both sides of the mismatch in the newly synthesized DNA strand. Mismatch removal then occurs by processive 5'–to–3' hydrolytic activity of MutSα-activated Exo1, which is

loaded at MutLα-catalyzed 5' strand breaks or at the 5' ends of Okazaki fragments on the lagging strand of DNA replication (reviewed in Li, 2008; Iyer *et al*, 2014).

MMR proteins play a role in both somatic and intergenerational CTG,·CAG, CGG and GAA repeat instability (van den Broek *et al*, 2002; Lin *et al*, 2006; Ku *et al*, 2010; Lokanga *et al*, 2014). Inactivation of *MSH2* and *MSH3*, but not *MSH6*, resulted in the stabilization or reduction of CAG and CTG repeats in HD and DM1 transgenic mice, respectively, suggesting that MutSα has a causative role and MutSβ has a preventive role in repeat instability (Manley *et al*, 1999; van den Broek *et al*, 2002; Savouret *et al*, 2003; Owen *et al*, 2005). Furthermore, PMS2 and MLH1 induced somatic instability in DM1 and HD transgenic mice, respectively (Gomes-Pereira *et al*, 2004b; Pinto *et al*, 2013). In FRDA, MMR proteins have been shown to recognise GAA expansion-containing triplexes and play a role in GAA repeat instability (Ku *et al*, 2010; Bourn *et al*, 2012; Ezzatizadeh *et al*, 2012; Du *et al*, 2012; Halabi *et al*, 2012; Hick *et al*, 2013; Ezzatizadeh *et al*, 2014). Deficit of any of the MSH2, MSH3, MSH6 or PMS2 increased GAA repeat intergenerational mutability (expansion and/or contraction) and loss of MSH2 or MSH6 led to a significant reduction in somatic GAA repeat expansions in FRDA transgenic mice. In contrast, loss of PMS2 promoted GAA repeat expansions in transmission from parents to offspring (Ezzatizadeh *et al*, 2012) and increased GAA repeat expansions in neuronal tissues of these mice, particularly the cerebellum and DRG (Bourn *et al*, 2012). MLH1 promoted intergenerational and somatic GAA repeat expansions (Ezzatizadeh *et al*, 2014). Furthermore, expression of the MutSβ complex, particularly the MSH3 subunit, was found to be necessary for expansion of GAA repeats in model cells and FRDA patient fibroblasts (Halabi *et al*, 2012), which led to the interesting possibility that inhibition of *MSH3* could slow down GAA repeat expansions in FRDA affected tissues and stop the progression of the disease.

1.3.2.1.2 Heterochromatinisation of the *FXN* gene

Epigenetic mechanisms which include DNA methylation, post-translational histone modifications and non-coding RNAs, regulate gene expression without involving changes in the primary DNA sequence. Dysregulation of chromatin organisation and function plays a pathogenic role in several neurodegenerative diseases for which chromatin-modifying compounds may be potentially therapeutic (reviewed in Jakovcevski & Akbarian, 2012).

1.3.2.1.2.1 Histone modifications

In eukaryotic cells, the nuclear DNA is packaged into a higher order structure, known as chromatin. The nucleosome is the fundamental unit of chromatin and consists of an octamer of four pairs of histone proteins H2A, H2B, H3, and H4 around which 147 bp of DNA are wrapped (Kornberg & Thonmas, 1974). Both unstructured histone tails and globular domains are subject to a vast array of post-translational modifications, namely methylation of arginine residues; methylation, acetylation, ubiquitination, ADP-ribosylation, and sumolation of lysines (K); and phosphorylation of serines and threonines (reviewed in Kouzarides, 2007).

Histone modifications can be involved in the establishment of global chromatin environments (euchromatin and heterochromatin), but also regulate DNA-related metabolic processes including transcription, recombination, DNA repair, replication, and chromosome condensation (Kouzarides, 2007). Modifications that are associated with active transcription, such as acetylation of histones H3 and H4 or di- or trimethylation of H3K4, are commonly referred to as euchromatin modifications. Heterochromatin modifications are associated to inactive genes or regions, such as H3K9me and H3K27me (reviewed in Li *et al*, 2007). Moreover, in actively transcribed genes, acetylation is distributed in distinct localised patterns in the promoter, 5' and 3' end of coding regions (Fig. 1.4); within the promoter there are two nucleosomes flanking the initiation site that are hypoacetylated at certain lysines and are enriched in the H2A variant Hzt1; the initiation site shows low nucleosome occupancy; lysine trimethylation is enriched in the coding region; and methylation of H3K4, H3K36, H3K79 indicates transcription initiation and elongation (Kouzarides, 2007; Li *et al*, 2007).

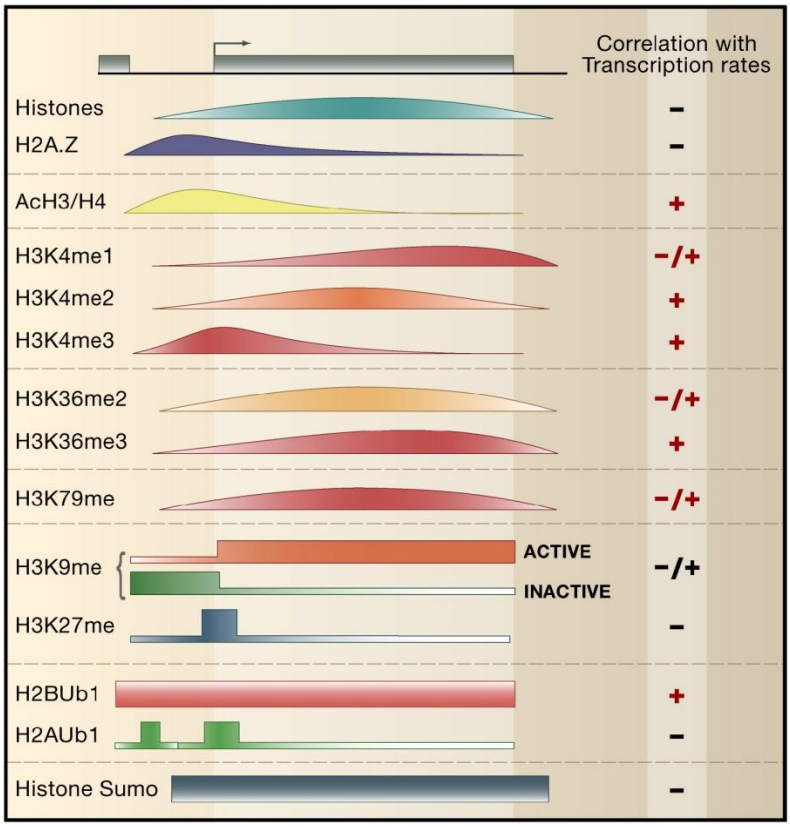

Figure 1.4 Distribution pattern of histone modifications involved in transcription regulation. The distribution of histones and their modifications are mapped on an arbitrary gene relative to its promoter (5′ end), ORF, and 3′ end. The curves represent the patterns that are determined via genome-wide approaches. The squares indicate that data are based on only a few case studies. With the exception of the data on K9 and K27 methylation, most of the data are based on yeast genes (Image adapted from Li *et al*, 2007).

It is still unclear why certain histone modifications have different functions. However, histone acetylation results in a change in the net charge of nucleosomes, which can loosen inter- or intranucleosomal DNA-histone interactions (Reinke & Hörz, 2003). Moreover, individual histone modifications or modification patterns can be recognized by other proteins that influence chromatin dynamics and function (Strahl & Allis, 2000; Jenuwein & Allis, 2001).

Histone acetylation occurs at multiple lysine residues due to a variety of histone acetyltransferase complexes (HATs) (Brown *et al*, 2000) and is removed by histone deacetylases (HDACs). Similarly, histone methylation is controlled by histone methyltransferases (HMTs) and histone demethylases (HDMs), which have been linked to transcriptional activation and repression (Kouzarides, 2007).

HATs utilize acetyl CoA as cofactor and catalyse the transfer of an acetyl group to the ε-amino group of lysine side chains, neutralising the lysine's positive charge and

potentially weakening the interactions between histones and DNA. HATs such as Gcn5, TAF1, and p300/CBP are recruited to the promoter upon transcription initiation (Sterner & Berger, 2000; Strahl & Allis, 2000) and levels of acetylation are proportional to the transcription rate (Pokholok *et al*, 2005). There are two different families: type-A HATs are predominantly cytoplasmic, acetylating free histones but not those already deposited into chromatin; type-B HATs acetylate newly synthesized histone H4 at K5 and K12 (as well as certain sites within H3), and this pattern of acetylation is important for deposition of the histones (Bannister & Kouzarides, 2011).

HDACs reverse lysine acetylation, which restores the positive charge of the lysine, leading to stabilisation of the local chromatin architecture and transcriptional repression. There are four classes of HDACs: classes I and II contain enzymes that are most closely related to yeast scRpd3 and scHda1, respectively; class III (referred to as sirtuins) are homologous to yeast scSir2 and require a specific cofactor for its activity, NAD^+; class IV has only a single member, HDAC11. A single HDAC is able to deacetylate multiple sites within histones (Bannister & Kouzarides, 2011).

HMTs do not alter the charge of histones and are divided into lysine-specific and arginine-specific groups. They have great specificity and usually modify only one particular histone residue. All histone lysine methyltransferase catalyse the transfer of a methyl group from S-adenosylmethionine to a ε-amino group of a lysine.

HDMs are classified into two distinct enzyme families: the nuclear amine oxidase homologs, like LSD1, and the JmjC-domain proteins. All HDMs are very specific to their lysine targets and sensitive to the degree of lysine methylation (Kouzarides, 2007).

There is an increasing amount of evidence supporting a link between epigenetic dysregulation and unstable expanded diseases, including FRAXA, FRDA, DM1, HD, SCAs and SBMA. Histone modification changes generally involve transition from active chromatin marks (H3K4me2/3, H3K9Ac) to those associated with silenced chromatin (H3K9me2/3; H4K20me3, H3K27me3). This is generally associated with decreasing transcription and protein levels in association with increasing antisense RNAs (Evans-Galea *et al*, 2013). For example, *FMR1* silencing occurs together with enrichment of H3K9me3 and H4K20me3 in exon 1 of the expanded *FMR1* allele, just downstream of the CGG expansion, prior to DNA methylation (Kumari & Usdin, 2010); expanded *ATXN8OS* shows increased H3K9me2 and reduced H3K14ac upstream of the CTG expansion (Chen *et al*, 2009); and in DM1, expanded CTG repeats induce the heterochromatinization of the *DMPK* locus (Otten & Tapscott, 1995).

Saveliev and colleagues provided the first report indicating that expanded GAA repeats could induce heterochromatin-mediated impairment of a gene (Saveliev *et al*, 2003). In this study, long GAA repeats fused to the hCD2 gene induced heterochromatin silencing even when the construct was integrated into a euchromatic environment, exerting a phenomenon similar to position effect variegation (PEV) and implying that expanded GAA repeats could nucleate heterochromatin. Moreover, GAA expansion-mediated silencing was sensitive to dosage of the classical PEV modifier heterochromatin protein 1 (HP1), raising the possibility that PEV modifiers could potentially play a role in FRDA (Saveliev *et al*, 2003).

Several studies have since shown that a similar epigenetic mechanism occurs at the endogenous *FXN* gene in FRDA. Chromatin immunoprecipitation (ChIP) analysis using primary lymphocytes and cell lines derived from FRDA patients (Herman *et al*, 2006; Greene *et al*, 2007; Punga & Bühler, 2010; Rai *et al*, 2010; Kim *et al*, 2011; Kumari *et al*, 2011; Groh *et al*, 2014) has shown that the first intron of expanded *FXN* alleles is enriched in heterochromatin marks, such as hypoacetylated histones H3 and H4, methylated histones H3K9me2 and H3K9me3, the latter providing the binding site for HP1 (Saveliev *et al*, 2003). These epigenetic changes are also detected in the brain, cerebellum, and heart from FRDA patients (Al-Mahdawi *et al*, 2008) and in the brain and heart of FRDA mouse models (Al-Mahdawi *et al*, 2008; Rai *et al*, 2008).

While ChIP data among these studies are widely accepted for the region flanking the GAA expanded repeat, there are contradictory reports regarding the chromatin landscape of the *FXN* promoter. Some studies show no significant enrichment in H3K9me2 and H3K9me3 or hypoacetylated histones on the *FXN* promoter of FRDA patient-derived lymphoblastoid cell lines compared to healthy controls (Punga & Bühler, 2010; Kim *et al*, 2011). Consistently, Punga & Bühler (2010) found similar levels of the initiating form of RNAPII (phosphorylated at serine 5 on its C-terminal domain) on the promoter and immediately upstream of the GAA repeats in lymphoblastoid-derived cells from healthy controls and FRDA patients. However, there is also evidence showing that histone marks typical of transcriptionally repressed genes can be found at the *FXN* promoter and change its accessibility to regulatory factors important for transcription initiation (Al-Mahdawi *et al*, 2008; De Biase *et al*, 2009; Kumari *et al*, 2011; Chutake *et al*, 2014; Groh *et al*, 2014). In line with this, expanded GAA repeats can interfere with both *FXN* transcription initiation and elongation (Kumari *et al*, 2011; Chutake *et al*, 2014; Groh *et al*, 2014).

1.3.2.1.2.2 DNA methylation

DNA methylation results from the conversion of cytosine to 5-methylcytosine (5mC) predominantly in the context of CpG islands. The position of DNA methylation in the transcriptional unit influences its relationship to gene control. In particular, methylation in the immediate vicinity of a TSS blocks initiation, but methylation in the gene body does not block and might even stimulate transcription elongation and have an impact on splicing. Methylation also has a role in genome stability: it is important for chromosomal stability in repeat regions, such as centromeres, and is also likely to suppress the expression of transposable elements (reviewed in Jones, 2012).

When genes with CpG islands at their TSS are active, their promoters are usually characterized by nucleosome-depleted regions at the TSS, which are often flanked by nucleosomes containing the histone variant H2A.Z and are marked with H3K4me3 (Kelly *et al*, 2010). Promoters containing CpG islands can be repressed by various mechanisms, such as repression mediated by Polycomb proteins. Methylated CpG islands at TSSs are associated with silent genes and cannot initiate transcription after the DNA has been assembled into nucleosomes. While certain forms of histone methylation cause local formation of heterochromatin, which is readily reversible, DNA methylation leads to a stable long-term repression (reviewed in Cedar & Bergman, 2009). Promoters containing CpG islands that are already silenced by Polycomb complexes are much more likely than other genes to become methylated, which suggests that the silent state precedes methylation. Methylation appears to serve as a 'lock' to reinforce a previously silenced gene state (Gal-Yam *et al*, 2008).

Intragenic CpG islands can become extensively methylated, however this does not block transcription elongation despite being marked by H3K9me3 and bound by methyl-CpG-binding protein 2 (MECP2), which are chromatin features that are associated with repressed transcription when they are present at the TSS. Thus, in mammals, it is the initiation of transcription but not transcription elongation that seems to be sensitive to DNA methylation silencing (Nguyen *et al*, 2001). It is thought that intragenic methylation is a mechanism for silencing repetitive DNA elements, such as retroviruses, LINE1 elements, Alu elements and others. Methylation blocks initiation of transcription at these elements while at the same time allowing transcription of the host gene to run through them (Yoder *et al*, 1997).

In mammals, DNA methyltransferase (DNMT) enzymes include three functional proteins: DNMT1 preferentially methylates hemi-methylated DNA and is responsible for

methylation during DNA replication, while DNMT3a and DNMT3b have an equal preference for hemi-methylated and non-methylated DNA and so have been classified as *de novo* methyltransferases. Active demethylation requires a mechanism that ultimately involves cell division or DNA repair and the excision of the base rather than the removal of the methyl group directly from the 5mC. This involves enzymes such as the ten-eleven translocation (TET) methylcytosine dioxygenases, activation-induced cytidine deaminase (AID) and thymine DNA glycosylase (TDG) in active and passive demethylation and in gene activation (reviewed in Jones, 2012).

When the expansion length inscreases, distinct but variable changes in CpG methylation state occur in the upstream and/or downstream region of the expansion and also, in some cases, in the promoter region of mutated genes, causing unstable expansion diseases. In FRAXA, the CGG expansion triggers widespread methylation of the *FMR1* gene, where most CpG sites within the repeat are methylated, as are those in the promoter region upstream of the repeat. DNA methylation levels correlate with *FMR1* expression (Evans-Galea *et al*, 2013). In DM1, altered DNA methylation of *DMPK* has been identified, with upstream elements being highly methylated and downstream elements being unmethylated (López Castel *et al*, 2011).

Initial studies of DNA methylation within the *FXN* gene showed that expanded GAA repeat induced hypermethylation of specific CpG sites in the upstream region of the GAA repeat sequence in FRDA patient-derived lymphoblastoid cells (Greene *et al*, 2007). These results were later confirmed in the brain, heart, and cerebellum tissues from FRDA patients and in tissues of FRDA YAC transgenic mice (Al-Mahdawi *et al*, 2008). In line with these findings, two studies on peripheral blood mononuclear cells also reported significantly higher levels of DNA methylation upstream of the expanded repeats on the *FXN* locus, (Castaldo *et al*, 2008; Evans-Galea *et al*, 2012). Both found a negative correlation between the level of DNA methylation in this region and the age of disease onset. Castaldo and colleagues also reported that the degree of methylation is proportional to the length of GAA1, the smallest of the two expanded *FXN* alleles (Castaldo *et al*, 2008), and Evans-Galea and colleagues suggested a positive correlation between *FXN* gene silencing, disease severity and the level of DNA methylation (Evans-Galea *et al*, 2012). More recently, a new analysis able to distinguish between 5-hydroxymethylcytosine (5hmC) and 5mC revealed that, for at least one CpG site within the upstream GAA repeat region, the increased level of DNA methylation predominantly comprises 5hmC rather than 5mC (Al-Mahdawi *et al*, 2013).

1.3.2.1.2.3 RNA interference (RNAi)

Antisense transcripts are common in the genome and may play important roles in RNA stability and transcriptional activity. Antisense transcription can generate double-stranded RNA, which trigger RNAi, enabling post-transcriptional repression of specific mRNAs. Double stranded RNAs, such as microRNAs, are processed by the ribonuclease Dicer into 21 nucleotide fragments. These short RNAs are subsequently loaded in the RNA-induced silencing complex (RISC), which targets mRNAs for degradation or inhibits their translation (Dion & Wilson, 2009).

Antisense transcription also plays a role in the pathogenesis of several unstable expanded diseases. In FRAXA, *FMR1* silencing is mediated by the *FMR1* mRNA, which hybridizes to the complementary CGG-repeat portion of the *FMR1* gene to form an RNA/DNA duplex (Colak *et al*, 2014). At the *DMPK* locus, high levels of antisense transcription are associated with regional H3K9me2, HP1γ recruitment and loss of CTCF binding (Cho *et al*, 2005).

Besides heterochromatic histone marks, the *FXN* promoter of FRDA fibroblasts shows elevated levels of HP1, increased levels of a frataxin antisense transcript 1 (FAST1) together with depletion of CTCF binding at the 5' UTR region of the *FXN* gene (De Biase *et al*, 2009). Furthermore, increased expression of FAST1 have also been shown in FRDA mouse fibroblasts (Sandi *et al*, 2014b) and reduced CTCF binding at the 5' UTR region of the *FXN* gene in FRDA cerebellum tissue (Al-Mahdawi *et al*, 2013). It is still unclear where the starting site of FAST1 transcription is located. However, FAST1 extends from the first exon and intron towards the *FXN* promoter (De Biase *et al*, 2009).

1.4 Frataxin protein

1.4.1 Normal frataxin function

The frataxin protein (reviewed in Pastore & Puccio, 2013; Martelli & Puccio, 2014) is a small, highly conserved acidic mitochondrial protein essential for life, as demonstrated in experiments using knockout animals where complete absence of frataxin led to early embryonic death (Cossée *et al*, 2000). It is highly expressed in the dorsal root ganglia, the granular layer of the cerebellum as well as the heart, pancreas, thymus, brown fat, muscle and liver (Campuzano *et al*, 1996; Koutnikova *et al*, 1997). Interestingly, Xia and colleagues suggested that tissue-specific expression of different frataxin isoforms, particularly the extra-mitochondrial isoforms, may explain the pathology in other FRDA-

affected tissues (Xia *et al*, 2012). In this study, two novel tissue-specific transcript variants were identified, encoding two isoforms of frataxin (isoform II and III) lacking the mitochondrial targeting sequence and were therefore different from the canonical transcript (encoding isoform I). Isoforms II and III were functional and were specifically localised in affected cerebellum and heart tissues, respectively. Functional assays revealed that isoform I and II protect iron-sulphur clusters (ISCs) from oxidative damage, and FXN III acts as a mitochondrial enhancer of ISC-biogenesis.

Sequence alignment of human, mouse, yeast (Yfh1), and bacterial (CyaY) frataxin showed two distinct regions. The N-terminal (70-90 amino acids) is poorly conserved among eukaryotes, whereas the C-terminal (100-120 amino acids) is highly conserved in most organisms and contains the functionally relevant part of the protein (Huynen *et al*, 2001). Although the protein is nuclear-encoded and synthesised as a precursor polypeptide, it is expressed in the cytoplasm and imported to the mitochondrial matrix through an import signal contained in the N-terminus (Koutnikova *et al*, 1997). To obtain the intermediate and mature forms of the human frataxin (Fig. 1.5), the 210-amino acid precursor (23 kDa) undergoes a two-step maturation process *in vivo* (Schmucker *et al*, 2008). The mitochondrial processing peptidase (MPP) initially cleaves the precursor between Gly41 and Leu42, generating the intermediate form (i-FXN; 19 kDa) and then performs a second cleavage to generate the mature form (m-FXN) (Koutnikova *et al*, 1998). The mature form was initially reported to result from cleavage between Ala55 and Ser56 (m_{56}-FXN; 17 kDa) (Cavadini *et al*, 2000). However, two additional studies have reported two different mature forms resulting from cleavage sites at residues Leu78-Arg79 (m_{78}-FXN; 14.5 kDa) and Lys80-Ser81 (m_{81}-FXN; 14.2 kDa) (Condò *et al*, 2007; Yoon *et al*, 2007). Finally, Schmucker and colleagues demonstrated that the major endogenous form of mature frataxin corresponds to the m_{81}-FXN and that the m_{56}-FXN form is not found in normal physiological conditions (Schmucker *et al*, 2008). Furthermore, m_{78}-FXN was shown to be a degradation product.

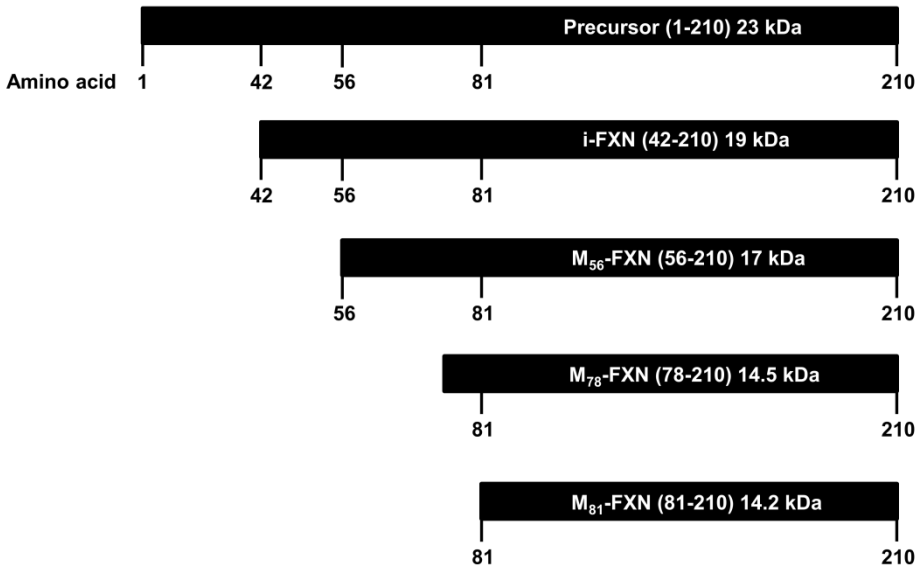

Figure 1.5 Maturation of the frataxin protein. To obtain the intermediate and mature forms of the human frataxin, the 210-amino acid precursor undergoes a two-step maturation process .The intermediate (i-FXN) and the three processed forms m_{56}-FXN, m_{78}-FXN and m_{81}-FXN are represented. The mature frataxin form corresponds to the m_{81}-FXN (adapted from Schmucker *et al*, 2008).

CyaY, Yfh1 and human frataxin are highly soluble monomeric proteins in the absence of iron. Under aerobic conditions and iron overload, oligomers are observed for both CyaY and Yfh1 (Cook *et al*, 2006; Adinolfi *et al*, 2009). Human frataxin assembles only under extreme conditions and in an iron-independent manner (O'Neill *et al*, 2005). Furthermore, formation of oligomers is not a requisite for m_{81}-FXN mature form to be functional (Schmucker *et al*, 2008).

The exact function of frataxin remains unclear, however frataxin deficiency in FRDA shows that it is primarily involved in the biosynthesis of ISCs (Fig. 1.6), the complexes that serve as prosthetic groups for a variety of enzymes involved in energy and iron metabolism, purine synthesis and DNA repair, making frataxin of vital importance for the synthesis of the mitochondrial respiratory chain complexes I, II and III enzymes and aconitase (Rotig *et al*, 1997).

Under normal conditions, frataxin binds iron directly and can function as an iron chaperone during heme and ISC assembly (Aloria *et al*, 2004). The first step of ISC biogenesis involves the assembly of inorganic iron and sulphur into an ISC on a scaffold protein, IscU (Isu1 in yeast). In this process, sulphur is provided through a persulphide intermediate by a cysteine desulphurase IscS (NFS1 and ISD11). Nfs1 is a cysteine desulphurase that converts cysteine to alanine and a highly reactive persulphide used in the

synthesis of sulphur bioorganic derivatives; and Isu1 is the transient scaffold protein on which the cluster assembles. The constitutive Nfs1 homodimer and two copies of monomeric Isu1 (or their bacterial orthologs IscS and IscU) form a binary complex (Pastore & Puccio, 2013). Frataxin interacts with several proteins of this pathway. It was shown that Yfh1 and human frataxin bind to the central ISC assembly complex (Gerber *et al*, 2003; Ramazzotti *et al*, 2004), regulating the interaction between the scaffold IscU and the desulfurase IscS and to modulate the formation of [Fe4S4] clusters (Colin *et al*, 2013). Moreover, frataxin may act as an iron donor which helps to solubilize and transport iron to the ISC biogenesis pathway (Yoon & Cowan, 2003).

Other functional interactions were suggested between frataxin and ferrochelatase, the enzyme that catalyzes the final step of heme biosynthesis by inserting the ferrous ion into porphyrin (Yoon & Cowan, 2004), mitochondrial aconitase (Bulteau *et al*, 2004), complex II of the mitochondrial respiratory chain (González-Cabo *et al*, 2005) and several chaperones (GRP75, Ssc1) (Shan *et al*, 2007).

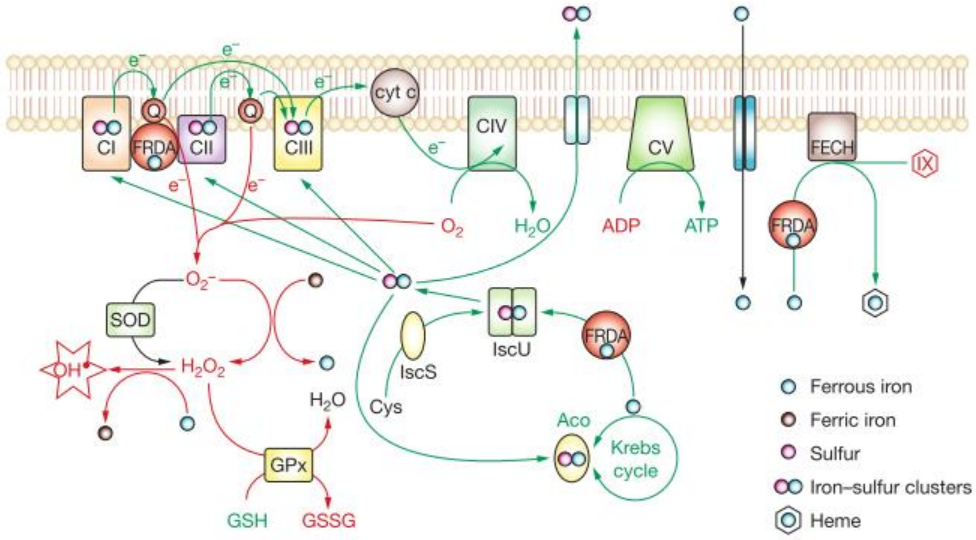

Figure 1.6 Frataxin function and oxidative stress in FRDA. Schematic representation of the postulated functions of frataxin, including ISC and heme biogenesis, and direct interaction with the respiratory chain complexes. Green arrows and text indicate molecules and pathways that have decreased activity in frataxin deficiency; red arrows and text indicate molecules and pathways that have increased activity in frataxin deficiency. Abbreviations: Aco, aconitase; C(I-V), respiratory chain complex I-V; Cys, cysteine; cyt c, cytochrome c; e–, electron; FECH, ferrochelatase; FRDA, frataxin; GPx, glutathione peroxidase; GSH, reduced glutathione; GSSG, oxidized glutathione; H_2O_2, hydrogen peroxide; IscS, cysteine desulfurase; IscU, ISC scaffold protein; IX, protoporphyrin IX; OH•, hydroxyl radical; Q, coenzyme Q; SOD, superoxide dismutase (Pandolfo, 2008).

1.4.2 Frataxin defficiency

Consequently to the defect in ISC-containing proteins, cells derived from FRDA patients show mitochondrial damage, lower ATP production, and impaired iron utilization, leading to mitochondrial iron accumulation (Ristow *et al*, 2000; Michael *et al*, 2006; Li *et al*, 2008) and formation of free radicals, which may further reduce the levels of oxidative-sensitive ISC proteins and cause increased cellular damage (Chantrel-Groussard *et al*, 2001). In particular, accumulation of large amount of iron inside mitochondria as granular iron deposits have been reported in a yeast strain deficient for the yeast frataxin homolog (Yfh1) (Babcock *et al*, 1997), conditional mouse model reproducing the cardiac phenotype (MCK mouse) (Puccio *et al*, 2001), cardiomyocytes of FRDA patients (Lamarche *et al*, 1980), and heart autopsies of individuals with FRDA (Michael *et al*, 2006).

The cerebellar dentate nucleus contains high levels of iron, but showed no difference in iron concentration in FRDA patients compared with healthy controls (Koeppen *et al*, 2007). However, Koeppen and colleagues observed a shift in expression of iron-related proteins such as transferrin receptor 1 (TFR1), ferritins (FRTs) and ferroportin (FPN), which suggests a change in iron metabolism (Koeppen *et al*, 2007). Furthermore, iron can relocate from neurons to microglia of the dentate nucleus in FRDA (Koeppen *et al*, 2012). Similarly, DRGs from FRDA patients do not show higher iron concentration than healthy controls, but the expression of the iron-storage FRT proteins increases in neighbouring cells (Koeppen *et al*, 2013).

1.5 Therapeutic strategies

There is currently no proven treatment for FRDA. Several clinical trials have been conducted with the most promising strategies in order to rescue the pathological deficits seen in FRDA patients. These potential therapies were primarily aimed at managing symptoms, and thus improving the quality of life, through rescue of the increased oxidative stress and accumulation of iron in the mitochondria. As the molecular mechanisms underlying FRDA pathogenesis become better understood, other strategies have been developed to increase *FXN* mRNA and protein levels.

1.5.1 Antioxidants

Treatment approaches to address mitochondrial pathogenesis include the use of enhancers of mitochondrial function, such antioxidants or iron chelators. Coenzyme Q_{10} is a potent antioxidant and electron carrier between the respiratory chain complexes I and II,

and from oxidation of fatty acids and branched chain amino acids, to complex III resulting in the ultimate production of ATP. Idebenone is a short-chain benzoquinone structurally analog to coenzyme Q_{10}. Initial results with these antioxidants were promising, resulting in improvement in cardiac and skeletal muscle bioenergetics (coenzyme Q_{10}) and reduction of cardiac hypertrophy (idebenone) (Schulz *et al*, 2009). However, larger phase III randomized, placebo-controlled trials failed to show disease modifying efficacy of idebenone treatment (Parkinson *et al*, 2013).

1.5.2 Iron quelators

To address the iron accumulation in mitochondria and depletion in the cytoplasm, the use of iron chelators was considered to act on the labile iron pools and subcellular compartments (such as the mitochondria) without depleting transferrin-bound iron from the plasma (Schulz *et al*, 2009). Deferiprone has been used to treat iron overload in individuals with hemoglobinopathies and can easily cross the blood–brain barrier and cellular membranes. In addition, because of its relatively low affinity for iron, it is not likely to cause iron depletion. These properties made deferiprone an interesting candidate as an FRDA therapy (Pandolfo & Hausmann, 2013). Deferiprone was tested in a pilot, open-label study and in a randomized, placebo-controlled phase II clinical trial. Although in the pilot study deferiprone reduced the iron content in the dentate nucleus of FRDA patients, the phase II study failed to demonstrate an improvement of ataxia, which was even worsened by high doses of the drug. However, deferiprone was able to reduce heart hypertrophy at all tested doses (Pandolfo & Hausmann, 2013).

1.5.3 Gene therapy

Gene therapy can potentially increase frataxin levels and alleviated FRDA symptoms. Recombinant adeno-associated viral (AAV) and lentiviral vectors expressing frataxin cDNA were shown to partially correct sensitivity to oxidative stress in FRDA primary fibroblasts (Fleming *et al*, 2005). Herpes simplex virus type 1 (HSV-1) amplicon vectors expressing the whole *FXN* genomic locus were used to transduce FRDA primary fibroblasts and were able to successfully rescue sensitivity to oxidative stress (Gomez-Sebastian *et al*, 2007). This same construct resulted in long-term persistent *FXN* expression in the brain of wild-type mice after injection into the adult mouse cerebellum (Gimenez-Cassina *et al*, 2011). Recently, a new study achieved not only correction, but also reversal, of a severe cardiomyopathy phenotype in a conditional FRDA mouse model following

intravenous injection of AAV constitutively expressing the human *FXN* gene (Perdomini *et al*, 2014). Importantly, *FXN* expression also restored compromised ventricular function to healthy levels (Perdomini *et al*, 2014).

Introduction of *FXN* into cells can also be achieved via non-viral-based vector delivery (reviewed in Evans-Galea *et al*, 2014). Bacterial and yeast artificial chromosomes (BACs and YACs, respectively) expressing human *FXN* were able to rescue the embryonic lethality of *FXN*-deficient mice (Pook *et al*, 2001; Sarsero *et al*, 2004). An exogenous replacement of frataxin, directed to the mitochondria via an N-terminal fusion of the transduction domain of the cationic peptide transactivator of the HIV transcription protein TAT, improved cardiac function and increased lifespan in a conditional frataxin knockout mouse (Vyas *et al*, 2012). Recently, a new study demonstrated a 1.6-fold increase in *FXN* mature mRNA and a 1.6–1.8-fold increase in frataxin protein, after nucleoinfection of human FRDA fibroblasts with a transcription activator-like effector protein (TALE) fused to a transcription activation domain (VP64) engineered to bind to the *FXN* promoter (Chapdelaine *et al*, 2013).

Zinc finger nucleases (ZFNs) have been shown to specifically recognise and cleave CAG or CTG repeat tracts. When a single ZFN was expressed in the target cells, repeat contractions could readily be detected, indicating that CTG and CAG duplexes were formed in the cells (Liu *et al*, 2010). In FRDA, Li and colleagues successfully corrected human FRDA cells via ZFN-mediated excision of expanded GAA repeats (Li *et al*, 2015).

1.5.4 Epigenetic therapies

Epigenetic strategies have been used in unstable expanded diseases, for example treatment of FRAXA patients lymphoblastoid cells with 5-aza-2'-deoxycytidine, a DNMT inhibitor, either alone (Chiurazzi *et al*, 1998) or in combination with HDAC inhibitors (Chiurazzi *et al*, 1999), efficiently reverses the *FMR1* promoter hypermethylation and restores mRNA and protein levels to normal. Moreover, HDAC inhibitors have been shown to restore the dysregulated gene expression in HD mice and ameliorate their neurodegenerative phenotype (Ferrante *et al*, 2003; Jia *et al*, 2012).

In pre-clinical studies, a family of HDAC inhibitors has been shown to increase *FXN* mRNA and protein levels in cells from FRDA patients and in FRDA mouse models (Gottesfeld *et al*, 2013). Initially, Herman and colleagues treated FRDA lymphoblastoid cells using a selection of commercially available HDAC inhibitors and showed that the benzamide compound BML-210 increased *FXN* mRNA levels and its analog, compound

4b, could directly increase acetylation of specific histone marks associated with the *FXN* gene to increase *FXN* mRNA levels in FRDA primary lymphocytes (Herman *et al*, 2006). Compound 4b was also shown to ameliorate the phenotype and transcriptional abnormalies in an HD transgenic mice (Thomas *et al*, 2008). Three 2-aminobenzamide compounds, named 106, 136, and 109, were synthesised from 4b to improve their *FXN*-upregulating action in FRDA cells and minimise their effect on unaffected control cells. Compound 106 has been shown to be an inhibitor of class I HDACs, with a preference for inhibition of HDAC3 (Xu *et al*, 2009). Compound 109 was shown to be the most promising compound after a long-term treatment using YG8R YAC transgenic mice (Sandi *et al*, 2011) as well as in neuronal cells derived from FRDA- induced pluripotent stem cells (iPSCs) (Soragni *et al*, 2014). This compound was taken into a Phase Ib clinical trial in FRDA patients, where drug treatment led to increases in *FXN* mRNA and histone acetylation at the *FXN* gene in peripheral blood mononuclear cells (Soragni *et al*, 2015). Similarly, the sirtuin protein deacetylase inhibitor nicotinamide (vitamin B3) also increases *FXN* mRNA levels in FRDA models (Chan *et al*, 2013), and recently in circulating lymphocytes from nicotinamide-treated FRDA patients in an exploratory, open-label, dose-escalation clinical trial (Libri *et al*, 2014).

1.6 Aims of this Thesis

In FRDA, expanded GAA repeats in intron 1 of the *FXN* gene reduce steady-state *FXN* mRNA levels through a poorly understood mechanism. In order to better understand the underlying pathogenesis and ultimately to design effective therapies for FRDA, it is important to generate cellular models that recapitulate the repressive hallmarks of the disease while providing efficient ways to quantify *FXN* expression. Therefore, the purpose of this study is to dissect the silencing mechanism of FRDA using genomic reporter models of FRDA and physiologically relevant carrier and FRDA patient-derived cells.

This thesis has been structured under four main aims:

1. Characterisation of the epigenetic silencing induced by expanded GAA repeats in a human cell model of FRDA (Chapter 3);
2. Investigation of the link between *FXN* localisation and expression at the single-cell level in fixed and living cells using the MS2 system (Chapter 4);

3. Analysis of *FXN* localisation in its native genomic environment and assessment of *FXN* association with specific nuclear subdomains in carrier-derived lymphoblastoid cells (Chapter 5);
4. Elucidation of the interplay between the repressive environment at the nuclear lamina (NL) and *FXN* repression in FRDA patient-derived cells (Chapter 5).

CHAPTER 2

Materials and methods

2.1 Culture of human cells

2.1.1 Cell culture

Epstein Barr virus–transformed lymphoblastoid cell lines GM14519 (from an FRDA clinically unaffected carrier, alleles with 1285+50 and 9 GAA repeats), GM15851 (from an unaffected individual with normal range of GAA repeats), GM16209 (from an FRDA clinically affected individual, both alleles with 800 GAA repeats) and GM15850 (from an FRDA clinically affected individual, alleles with 1030 and 650 GAA repeats) were obtained from the Human Genetic Cell Repository of the Coriell Institute (USA) and propagated in RPMI1640 medium supplemented with 15% Fetal Bovine Serum (FBS) and 2 mM L-glutamine. HEK FRT cells were cultured in Dulbecco's modified Eagle's medium (DMEM) medium supplemented with 10% FBS, 2 mM L-glutamine, 100 U/ml penicillin/streptomycin and 100 µg/ml Zeocin (Life Technologies). *FXN-GAA-Luc*, *FXN-Luc*, *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* clonal cell lines were propagated in complete DMEM medium (see above) supplemented with 100 µg/ml Hygromycin B (Life Technologies).

2.1.2 Routine subculture

Cell culture was carried out under aseptic conditions in a Class II tissue culture cabinet. Cells were maintained in saturated humidity incubators at 37 °C with 5% CO_2. Routine subculture of adherent cells was performed every 3 days. The growth medium was discarded and cells were washed with 10 mL phosphate-buffered saline (PBS). Cells were trypsinised using 2 mL trypsin-EDTA (Sigma) at 37 °C for 5 minutes. Trypsinisation was halted using 8 mL of normal growth medium and clumps of cells were broken up by repeated pipetting. Cell suspensions were diluted to the appropriate seeding concentration in fresh growth medium and distributed into new 75 cm^2 flasks. Cells were allowed to reach 80% confluency until the next subculture. Suspension cells were mixed to disperse any clumps by pipetting the cell suspension up and down. Cell viability was monitored using trypan blue exclusion and cells were counted with a haemocytometer: 10 µL of cell suspension was loaded into the chamber and cells counted in eight 4 x 4 mm squares; the number of cells per mL of cell suspension was calculated by multiplying the average number of cells by 10,000. Cell suspensions were diluted to a final concentration of $2x10^5$ cells/mL in 20 mL of fresh growth medium and distributed into new 25 cm^2 flasks. Flasks were incubated in an upright position until the next subculture.

2.1.3 Cryopreservation of cells

Adherent cells growing in a 75 cm^2 flask were trypsinised as described above, resuspended in 10 mL of normal growth medium, transferred to a 15 mL conical centrifuge tube and centrifuged at 900 x *g* for 10 minutes. The supernatant was aspirated and the cell pellet was resuspended in 1.5 mL of normal growth medium and 1.5 mL of 2x freeze mix (60% DMEM, 30% FBS, 10% dimethyl sulphoxide (DMSO)). One millilitre of cells was transferred to a previously labelled 1.8 mL cryovial (Nunc). Suspension cells were counted as described above. Five million viable cells/mL were centrifuged at 100 x *g* for 10 minutes. The cell pellet was resuspended in 1 mL of cold freeze medium (RPMI1640 medium with 30% FBS and 5% DMSO) and transferred to a cryovial. Cryovials were stored in a polystyrene box and allowed to freeze to -80 °C overnight, then transferred to a liquid nitrogen storage tank.

2.1.4 Thawing frozen cells

Cells were rapidly warmed in a 37 °C water bath. Thawed cells were diluted in 10 mL of normal growth medium in a 15 mL conical centrifuge tube and centrifuged at 900 x *g* for 10 minutes. The supernatant was discarded and cells were resuspended in 8 mL of fresh growth medium. Cells were then distributed into 25 cm^2 flasks at different dilution ratios and allowed to grow as described above.

2.1.5 Drug treatment

Lymphoblastoid cell lines GM15850 and GM15851 were treated with DMSO or C5 at a concentration of 20 µM for 48 hours before ChIP analysis. Actinomycin D (ActD) time-course was performed by adding DMSO or 8 µg/ml ActD to the media of *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* clonal cell lines before lysing cells every 4 hours up to 24 hours for RNA extraction. *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells were treated with DMSO, nicotinamide (10 mM) or compound 106 (10 µM) for 48 hours before RNA extraction and Immuno-FISH.

2.2 Bacterial cell culture

Unless otherwise stated, all bacterial work was performed using ElectroMax DH10B *E. coli* cells (Life Technologies). Bacterial cultures were grown either in Luria broth (LB) or on LB agar.

2.2.1 Preparation of electrocompetent DH10B *E. coli*

Highly electrocompetent DH10B bacteria were used for all cloning experiments. A single bacterial colony was inoculated into 1.5 mL LB with the appropriate antibiotic and incubated overnight at 37 °C with shaking at 225 rpm. The culture was tipped into 100 mL LB with the appropriate antibiotics in a 1 L conical flask and incubated with shaking at 37 °C. After 1.5 hours, the optical density at 600 nm was determined using a cell density meter (WPA) every 15 minutes until it reached 0.35-0.36. The culture was then transferred to a 250 mL centrifuge flask and allowed to chill on ice at 4 °C for 40 minutes. Cells were pelleted at 5000 rpm for 15 minutes at 4 °C in an Avanti J-E centrifuge with a JLA-10.5 rotor. The supernatant was discarded and the pellet was washed three times with 100 mL ice-cold autoclaved 10% glycerol by centrifugation at 6000, 7000 and 7500 rpm for 15 minutes at 4 °C. After the final wash, the supernatant was carefully poured off and excess glycerol was removed with a clean tissue. The pellet was resuspended in the small residual amount of glycerol. Cells were distributed into 50 µL aliquots in microtubes and immediately snap-frozen in dry ice. Electrocompetent bacteria were stored at -80 °C.

2.2.2 Electroporation of electrocompetent cells

Bacterial transformation was achieved by electroporation using a Gene Pulser system (Bio-Rad). Electrocompetent bacteria (50 µL) were transformed in ice-cold 0.1 cm electroporation cuvettes according to the manufacturer's instructions. The following parameters were used: potential difference of 1800 V, capacitance of 25 µF and internal resistance of 200 Ω. Electroporated bacteria were incubated in 450 µL SOC medium (Life Technologies) for 1 hour at 37 °C with shaking, subsequently 2 µL and 100 µL were plated on LB agar petri dishes with the appropriate antibiotics and incubated overnight at 37 °C.

2.2.3 Transformation efficiency of electrocompetent cells

To measure the competence efficiency of electrocompetent cells, 1 µL of pUC19 plasmid DNA (10 pg/µL; Life Technologies) was electroporated into a 50 µL aliquot of electrocompetent cells as described above. Bacteria were plated on a LB agar petri dish with ampicillin (100 µg/mL) selection and allowed to grow overnight at 37 °C. Colonies were counted and the transfection efficiency was calculated as the number of colony forming units (CFU) per µg DNA using the following formula:

$$\frac{\text{CFU in control plate}}{\text{pg pUC19 used in transformation}} \times \frac{\text{total volume of culture}}{\text{volume plating medium}} \times \frac{1x10^6\text{pg}}{\mu\text{g}}$$

High transformation efficiencies of >1 x 10^9 transformants/μg pUC19 were routinely obtained using the described method.

2.3 DNA manipulation

2.3.1 Miniprep

A 1.5 mL LB culture with the appropriate antibiotic was seeded with a single bacterial colony and was incubated overnight at 37 °C with shaking. The culture was then tipped into a 1.5 mL microtube and cells were pelleted at 5000 rpm for 10 minutes. The supernatant was discarded and cells were resuspended in 70 μL of STET resuspension buffer (8% sucrose, 5% Triton X-100, 50 mM EDTA and 50 mM Tris-Cl at pH8). Cells were lysed with 200 μL of alkaline SDS solution (1% SDS, 0.2 M NaOH) and immediately neutralized with 150 μL of 7.5 M ammonium acetate. Microtubes were inverted three times and placed on ice for 5 minutes. The neutralised lysate was centrifuged at 13000 rpm for 20 minutes at 4 °C and the supernatant was transferred to a new microtube. DNA was precipitated by adding 250 μL isopropanol and was pelleted by centrifugation at 10000 rpm for 8 minutes. The supernatant was discarded and the pellet washed with 200 μL of 70% ethanol and then centrifuged at 10000 rpm for 8 minutes. The ethanol wash was aspirated and the DNA pellet allowed to air dry for 5-10 minutes. DNA was resuspended in 50 μL of Tris-EDTA (TE) containing 5 μg/mL RNase A.

2.3.2 Maxiprep

To obtain larger amounts of BAC DNA, 1.5 mL LB starter cultures seeded with a single bacterial colony were grown for 8 hours at 37°C with shaking and then were tipped into 250 mL LB with appropriate antibiotics in a 1 L conical flask and incubated with shaking for 16 hours at 37°C. Plasmid DNA was extracted using the Plasmid Maxi kit (Qiagen) with minor modifications to enhance DNA yields. Bacterial cells were pelleted by centrifugation at 6000 rpm for 15 minutes at 4 °C in a 250 mL centrifuge bottle using a JLA-10.5 rotor in an Avanti J-E centrifuge (Beckman Coulter). The pellet was resuspended in 15 mL of chilled Buffer P1 by shaking at 37 °C for 10 minutes. Cells were lysed with 15 mL of Buffer P2 with gentle mixing, and allowed to lyse for 5 minutes, swirling every 2.5 minutes. The reaction was neutralised with 15 mL of chilled Buffer P3, mixed by repeated inversion and incubated on ice for 20 minutes, swirling every 10 minutes. The lysate was centrifuged at 9000 rpm for 35 minutes at 4 °C. Qiagen-tip 100 columns were equilibrated

with 15 mL of Buffer QBT. The plasmid DNA supernatant was then filtered through a folded Kimtech 100 tissue (Sigma) placed inside equilibrated columns. Columns were washed twice with 30 mL of Buffer QC, and DNA was eluted into an Oakridge tube using 15 mL of Buffer QF pre-warmed to 55 °C. DNA was precipitated with 10.5 mL of isopropanol and pelleted by centrifugation at 14000 rpm for 30 minutes at 4 °C, using a JA-17 rotor. The pellet was then washed with 3.5mL of 70% ethanol and centrifuged again at 14000 rpm for 30 minutes at 4 °C. The pellet was allowed to air dry for 10 minutes before resuspension in 250 µL of TE overnight at 4 °C. The tube was centrifuged at 1000 rpm for 5 minutes to collect drops and transferred to a clean microtube. DNA was quantified using a NanoDrop spectrophotometer (Thermo Scientific).

2.3.3 Genomic DNA extraction from human cells

Genomic DNA from *FXN-GAA-Luc* and *FXN-Luc* clonal cell lines was isolated by standard phenol/chloroform extraction and ethanol precipitation. Cells were lysed in 500 µL of lysis buffer (0.6% SDS, 100 mM NaCl, 50 mM Tris-HCl pH 8, 20 mM EDTA) containing 50 µg/mL RNAse A with gentle shaking for 20 minutes at room temperature. The cell lysate was scraped into a 2 mL microtube and incubated overnight at 37 °C. The lysate was then incubated with 100 µg/mL of proteinase K overnight at 37 °C. The lysate was extracted twice with 500 µL phenol:chloroform (1:1) and twice with 500 µL of chloroform, using 1.5 mL phase-lock gel tubes (Eppendorf), mixed by inversion and centrifuged at 10000 rpm for 6 minute. After extraction, the lysate was transferred to a new microtube. DNA was precipitated with 30 µL of 5 M NaCl and 1.3 mL of absolute ethanol. After gently mixing by inversion, the precipitate was chilled on ice for 5 minutes and pelleted by centrifugation at 13000 rpm for 30 minutes at 4 °C. The DNA pellet was washed with 500 µL of 70% ethanol and centrifuged at 13000 rpm for 20 minutes at 4 °C. The pellet was allowed to air-dry briefly and resuspended in 100 µL of TE overnight at 4°C.

Genomic DNA from *FXN-GAA-MS2-Luc* and *FXN-MS2-Luc* clonal cell lines was isolated using Illustra Tissue and Cells GenomicPrep Mini Spin Kit (GE Healthcare), according to the manufacturer's instructions. DNA was quantified using a NanoDrop spectrophotometer (Thermo Scientific).

2.3.4 Polymerase Chain Reaction (PCR)

2.3.4.1 Standard PCR

Amplification of DNA fragments for cloning (section 2.3.9), sequencing (section 2.3.10), homologous recombination (section 2.3.11) and vector integration assessment was performed by standard PCR analysis. The integration of pBAC-*FXN-Luc*, p*BAC-FXN-GAA-Luc*, pBAC-*FXN-MS2-Luc* and pBAC-*FXN-GAA-MS2-Luc* at the docking site was assessed using primers pSV40-F (5'-CCAGTTCCGCCCATTCTC-3') and Hygro-R (5'-CAGCTATTTACCCGCAGGAC-3') and AmpliTaq Gold (Roche). DNA was denatured at 95 ºC for 10 minutes, followed by 35 cycles of denaturation at 95 ºC for 30 seconds, primer annealing at 58 ºC for 30 seconds and extension at 72 ºC for 1 minute, and then a final extension at 72 ºC for 8 minutes. For some reactions, the annealing temperature was adjusted to provide optimal specificity. Each reaction was carried out in 15 µl containing 1 µl of DNA (50 ng of genomic DNA or 10 ng of BACs), 1x Buffer, 1.5 mM $MgCl_2$, 0.4 mM deoxynucleotide triphosphate [dNTP] (Life Tecnologies), 0.05 µM of each primer, 1x Q Solution (Qiagen) and 1 U of AmpliTaq Gold (Roche). PCR products were visualised on a 1% agarose gel.

2.3.4.2 Amplification of GAA repeats

PCR amplification of the GAA repeat sequence was carried out on 200 ng of genomic DNA using primers 147F and 602R (Montermini *et al*, 1997a) and Expand Long Template DNA DNA Polymerase (Roche) as previously described (Campuzano *et al*, 1996; Al-Mahdawi *et al*, 2004). DNA was denatured at 94 ºC for 2 minutes, followed by 10 cycles of denaturation at 94 ºC for 10 seconds, primer annealing at 60 ºC for 30 seconds and extension at 68 ºC for 45 seconds, and then 20 cycles of denaturation at 94 ºC for 10 seconds, primer annealing at 60 ºC for 30 seconds and extension at 68 ºC for 1 minute with 20 seconds increments, and a final extension at 68ºC for 10 minutes. Each reaction was carried out in 25 µL containing 1 µL of DNA (200 ng), 1x Buffer 2, 125 µM of each dNTP (Life Tecnologies), 0.5 µM of each primer, 1x Q Solution (Qiagen) and 1.5 U of Expand Long Template DNA Polymerase (Roche). PCR products were visualised on a 1% agarose gel. The GAA repeat sequences were sequenced and digested with *Mbo*II (Holloway *et al*, 2011) to check for interruptions.

2.3.4.3 qPCR

For copy number determination, genomic DNA from *FXN-Luc*, *FXN-GAA-Luc*, *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* clonal cell lines was isolated using Illustra Tissue and Cells GenomicPrep Mini Spin Kit (GE Healthcare), according to the manufacturer's instructions. The number of transgene copies was determined by real-time PCR, using the relative standard curve method. Five-fold dilution standards were prepared to generate a standard curve for each primer pair. The upstream region of the GAA repeats in intron 1 of the *FXN* gene was amplified using the primers UpGAA-F and UpGAA-R (Herman *et al*, 2006) and normalising data by *GAPDH*, using the primers GAPDH-F and GAPDH-R (Herman *et al*, 2006). Three independent genomic DNA samples per cell line were quantified in triplicate by real-time PCR using the SYBR Green PCR Master Mix (Life Technologies). Each 25 μL reaction contained 2 μL of genomic DNA dilution, 1x SYBR Green PCR Master Mix and 70 nM of each primer. The assay was performed using the StepOnePlus Real-Time PCR system (Life Technologies) with the following protocol: 10 minutes at 95 °C for enzyme activation, followed by 40 cycles of denaturation at 95 °C for 15 seconds and primer annealing and extension at 60 °C for 1 minute. Specificity of amplification was monitored with a final dissociation stage which generates a melting curve. The number of transgene copies was determined by comparison with the acceptor cell line HEK FRT as reference sample, which carries three endogenous *FXN* loci as determined by FISH.

2.3.4.4 Small-pool PCR

Analysis of GAA repeat instability in the *FXN-GAA-Luc* transgene was performed as previously described (Gomes-Pereira *et al*, 2004a). Eight independent reactions were set up for each genomic *FXN-GAA-Luc* DNA concentration (6-600 pg) in siliconized microfuge tubes. PCR amplification of the GAA repeat sequence was performed using primers 147F and 602R and Expand Long Template DNA DNA Polymerase (Roche) as described in section 2.3.4.2. PCR products were resolved by electrophoresis on 1% agarose gels and bands were detected by Southern blotting using DIG High Prime DNA Labeling and Detection Starter Kit II (Roche), according to the manufacturer's instructions, with a digoxigenin-labelled $(TTC)_{10}$ probe. After agarose gel electrophoresis, DNA was denatured and transferred onto a hybond-N^+ membrane (GE Healthcare). DNA was allowed to hybridise to a digoxigenin-labelled $(TTC)_{10}$ probe at 55 °C overnight. Stringency washes were done with 2x SSC containing 0.1% SDS at room temperature for 5

minutes and 0.5x SSC containing 0.1% SDS at 60°C for 15 minutes. The membrane was blocked in 1x Blocking Solution for 30 minutes before incubation with an anti-digoxigenin antibody (1/10000 dilution) for 30 minutes. Chemiluminescent detection was performed using the CSPD ready-to-use solution.

2.3.4.5 Colony PCR

Colony PCR was performed to quickly identify which clones had taken up the plasmid of interest during homologous recombination. Individual colonies were picked from a petri dish into 50 µL of Milli-Q water in a 96-well plate. Bacteria were lysed by heating to 95 °C for 10 minutes and PCR analysis was performed using the AmpliTaq Gold (Roche) using the following conditions: DNA was denatured at 95 ºC for 10 minutes, followed by 35 cycles of denaturation at 95 ºC for 30 seconds, primer annealing at 58 ºC for 30 seconds and extension at 72 ºC for 1 minute, and then a final extension at 72 ºC for 8 minutes. PCR products were visualised on a 1% agarose gel.

2.3.5 Restriction digestion of plasmid DNA

Restriction enzymes, buffers and bovine serum albumin (BSA) were purchased from New England Biolabs (NEB). A restriction digestion reaction (15 µL) was prepared by adding the following reagents to a microfuge tube, on ice: 10 µL DNA from a miniprep (or 200 ng), 1.5 µL 10x restriction buffer, 0.75 µL 20x BSA, 0.3 µL restriction enzyme (3 units) and Milli-Q water up to 15 µL. The reaction was incubated at the required temperature for 2.5 hours. Small plasmids (<20 kb) were analysed by agarose gel electrophoresis (section 2.3.6), whereas BACs were analysed by pulsed-field gel electrophoresis (section 2.3.7).

2.3.6 Agarose gel electrophoresis

Agarose gels were prepared in 1x Tris-borate EDTA (TBE) (Sigma) at a concentration of 0.8-2% (w/v). The mix was heated in a microwave until the agarose was dissolved. The solution was cooled before the addition of ethidium bromide to a final concentration of 0.5 mg/mL. The agarose mix was poured into a gel tray with well combs in place. Electrophoresis was performed using a Sub-Cell GT gel electrophoresis system (Bio-Rad) with a covering volume of 1x TBE. Sample 6x loading buffer (1.5 g Ficoll 400, Orange G dye and Milli-Q water up to 10 mL) was added to the samples prior to loading.

A DNA ladder (Life Technologies) was routinely used as a size indicator. The gel was photographed using a GelDoc XR transilluminator system (Bio-Rad).

2.3.7 Pulsed-field gel electrophoresis

Pulsed-field electrophoresis was used to separate DNA fragments larger than 20 kb. Samples were run on a 1% agarose gel in 0.5x TBE buffer without ethidium bromide in a Bio-Rad CHEF DRII gel running tank containing 2 L of 0.5x TBE buffer. Running conditions were: 2 seconds of initial pulse time, 16 seconds of final pulse time, linear ramping, voltage of 6V/cm, buffer temperature at 14 °C, running time of 16 hours. The Mid-Range PFG II marker (NEB) and 1kb DNA ladder (Life Technologies) were used as size markers. Following electrophoresis, the gel was stained in 500 mL of 0.5x TBE buffer containing 1 mg/mL of ethidium bromide for 1 hour and photographed using the GelDoc XR system.

2.3.8 Purification of DNA fragments from agarose gels

Agarose gels were prepared in 1x TBE (Sigma) at a concentration of 1% (w/v) and electrophoresis performed as described in section 2.3.6. The band of interest was excised with the help of a handheld long wave UV illuminator. DNA was purified using the GeneClean kit (QBIO-gene) according to the manufacturer's instructions.

2.3.9 DNA ligation

Ligations were performed at 16 °C for 3 hours. Each 10 µL reaction contained: purified PCR product (in a vector: insert ratio of 1:1, 1:3 or 1:5), 1 µL purified linearised vector DNA, 1 µL of 10x T4 DNA ligase buffer, 1 µL of T4 DNA ligase (NEB) and Milli-Q water up to 10 µL. A negative control reaction contained all components except the insert. One microlitre of the ligation reaction was electroporated into electrocompetent DH10B *E.coli* cells, as described in section 2.2.2. Successful cloning was confirmed by colony-PCR, restriction enzyme digestions and sequencing.

2.3.10 DNA sequencing

Sequencing was performed on all plasmids after sequence modifications were completed. The region to be sequenced was first amplified by PCR using the standard PCR conditions (section 2.3.4.1) or the Expand Long Template protocol for GAA repeats (section 2.3.4.2). The PCR product was analysed by agarose gel electrophoresis prior to

sequencing. For purification, PCR buffer, salts and dNTPs were removed using exonuclease I (NEB) and shrimp alkaline phosphatase (SAP) (Promega). The ExoSAP reaction consisted of the following: 3.0 μL PCR product, 1.0 μL 10x SAP buffer, 1.0 μL SAP (500U/μL), 0.1 μL Exonuclease I (20U/μL) and 0.9 μL Milli-Q water. Samples were incubated at 37 °C for 1 hour and 80 °C for 20 minutes. Sequencing reactions were performed according to the BigDye Terminator v3.1 Cycle Sequencing protocol (Applied Biosystems). The 10 μL reaction mix consisted of the following: 5.0 μL PCR product (or 4.0 μL when sequencing GAA repeats), 1.0 μL Primer (10 μM), 1.5 μL 5x BigDye buffer, 0.5 μL BigDye (or 2.0 μL for GAA repeats) and 1.5 μL Milli-Q water. The sequencing reaction was performed using the following program: 96 °C for 1 minute, 30 cycles of 96 °C for 10 seconds, 50 °C for 5 seconds and 60 °C for 4 minutes. Samples were kept at 4 °C until DNA precipitation. DNA was precipitated as follows: 2 μL of 125 mM EDTA and 2 μL of 3 M sodium acetate were added to each well, followed by addition of 50 μL of 100% ethanol. The wells were covered with caps and the mixture vortexed before centrifugation at 3000 x *g* for 30 minutes at 4 °C. The supernatant was discarded and 70 μL of 70% ethanol was added to each well. After centrifugation at 1650 x *g* for 15 minutes at 4 °C, the supernatant was discarded and the wells allowed to air dry for 10 minutes protected from light. Sequencing was performed on a 3700 DNA Analyzer (Applied Biosystems) sequencing platform.

2.3.11 Homologous DNA recombination in *E. coli*

The pBAC-*FXN-GAA-MS2-Luc* was generated from the pBAC-*FXN-GAA-Luc* vector which carries the whole 80 kb *FXN* locus with exons 1–5b of the *FXN* gene, an insertion of the luciferase gene in exon 5a and a ~310 GAA repeat expansion in intron 1. A selection/counter-selection homologous recombination protocol based on Red/ET recombination (GeneBridges) was used. Electrocompetent bacteria containing the pBAC-*FXN-GAA-Luc* were prepared and transformed with 1 ng of the plasmid pSC101-BAD-gbaA (as described in section 2.2.2), which expresses the Red/ET recombinase enzymes under the control of an arabinose-inducible promoter (Gene Bridges). Since this plasmid contains a temperature-sensitive origin of replication, bacteria containing the Red/ET plasmid were always grown at 30 °C. Induction of homologous recombination enzymes was achieved by the addition of L-arabinose during the preparation of electrocompetent bacteria, as described in section 2.2.1, with the exception that 100 mL of LB bacterial culture in the appropriate antibiotic were grown at 30 °C with shaking until the optical

density reached 0.10-0.13. At this point, 1 mL of 10% of L-arabinose (w/v) was added to the culture, which was then incubated at 37 °C until the optical density was in the range 0.30-0.35. The culture was transferred to a 250 mL centrifuge flask and used to prepare electrocompetent cells as described in section 2.2.1.

Recombination using the *rpsL-neo* and the pSC101-BAD plasmids was carried out in two steps in *E. coli* according to the manufacturer's instructions. In the first step, a PCR product containing the *rpsL-neo* cassette flanked by 55 bp homology arms to either side of exon 2 was used to insert the cassette. The resulting PCR product was purified using a QIAquick PCR purification kit according to manufacturer's protocol (Qiagen), digested with *Dpn*I enzyme for 2 hours at 37 °C to remove any remaining template DNA and purified again. Eight microlitres of purified product were then electroporated into 50 μL of electrocompetent host cells containing the pBAC-*FXN-GAA-Luc* and expressing the Red/ET recombination proteins. Cultures were plated on LB agar with chloramphenicol and kanamycin and colonies were allowed to form at 37 °C overnight. Streptomycin sensitivity was also tested. Correct insertion of the *rpsL-neo* cassette was analysed by colony-PCR, PCR and restriction enzyme digestion followed by pulsed-field gel electrophoresis. GAA repeat expansion length was assessed by colony-PCR and PCR. Before the beginning of the second step of homologous recombination, persistence of pRed/ET in successful recombinants was analysed by restriction enzyme digestion.

Replacement of the *rpsL-neo* cassette in the second step of recombination by 24 MS2 Binding Sites in exon 2 was performed as follows. A plasmid containing exon 2 with an engineered *BamH*I restriction site and flanking homology arms (147 bp upstream and 172 bp downstream of exon 2) was purchased from GeneArt (Life Tecnologies). The 24 MS2 Binding Sites were excised by *BamH*I/*Bgl*II digestion from pCR4-24XMS2SL-stable (Addgene plasmid 31865) (Bertrand *et al*, 1998) and inserted in exon 2 at the *BamH*I site (Exon2-MBS). The *rpsL-neo* cassette was replaced with a PCR product containing the Exon2-MBS sequence flanked by homology arms, generating pBAC-*FXN-GAA-MS2-Luc*. As previously, 8 μL of purified PCR product were electroporated into 50 μL of electrocompetent host cells, containing the pBAC-*FXN-rspLneo-Luc* and expressing the Red/ET recombination proteins. Cultures were plated on LB agar with chloramphenicol and streptomycin and incubated overnight at 37 °C. Correct insertion of the 24 MBS cassette was confirmed by colony-PCR, PCR, restriction enzyme digestion followed by pulsed-field gel electrophoresis, and sequencing. GAA repeat expansion length was assessed by colony-PCR and PCR.

To generate pBAC-*FXN-MS2-Luc*, the ~310 GAA repeats in intron 1 in pBAC-*FXN-GAA-MS2-Luc* were replaced by the *rpsL-neo* cassette amplified with primers carrying 58 bp homology arms to sequences immediately upstream and downstream of GAA repeats. The *rpsL-neo* was then replaced with a PCR product containing 6 GAA repeats amplified from the genomic DNA of GM15851 using GAA-F and GAA-R primers (Campuzano *et al*, 1996). Successful construction was confirmed by PCR, restriction enzyme digestion followed by pulsed-field gel electrophoresis and sequencing.

2.3.12 Cre-*loxP* recombination for retrofitting

Cre-*loxP* mediated retrofitting of pBAC-*FXN-MS2-Luc* and pBAC-*FXN-GAA-MS2-Luc* vectors to pH-FRT-Hy was performed to insert the Flp-In promoter-less hygromycin cassette necessary to generate stable human clonal cell lines and a *lacZ* gene. For each BAC, three 30 μL reaction mixtures were prepared on ice: for reaction 1, 2 μg of each BAC, 50 ng of pH-FRT-Hy, 3 μL of Cre buffer, 1 μL of Cre recombinase and Milli-Q water; for reaction 2, 1 μg of each BAC, 50 ng of pH-FRT-Hy, 3 μL of Cre buffer, 1 μL of Cre recombinase and Milli-Q water; for reaction 3, 1 μg of each BAC, 3 ng of pH-FRT-Hy, 3 μL of Cre buffer, 1 μL of Cre recombinase and Milli-Q water. Samples were thoroughly mixed and briefly centrifuged. Each mixture was then incubated at 37 °C for 1 hour, followed by inactivation of the enzyme at 70 °C for 5 minutes. After leaving samples at room temperature for 10 minutes, the reaction was drop-dialysed against 50 mL of Milli-Q water for 1 hour, using 0.25 μm membrane filters (Millipore). Each sample was electroporated into electrocompetent DH10B bacteria as described in section 2.2.2, and plated on a 15 cm LB agar petri dish containing the relevant antibiotics (chloramphenicol for the BAC and ampicillin for the pH-FRT-Hy construct). Successful recombinants were identified by restriction enzyme digestion and pulsed-field electrophoresis.

2.4 Delivery of DNA into human cells

2.4.1 Lipofection

DNA was delivered to HEK FRT cells using Lipofectamine and Plus reagent (Life Technologies). Briefly, $1x10^6$ cells were seeded per 6 cm dish in DMEM medium supplemented with 10% FBS and 100 U/ml penicillin/streptomycin (Life Tecnologies) and allowed to adhere overnight. For each 6 cm dish, the following two transfection mixes were prepared: (i) tube A, 3200 ng of pBAC-*FXN-MS2-Luc* or pBAC-*FXN-GAA-MS2-Luc*

(500 ng when transfecting a small plasmid), 250 µL of Opti-MEM medium and 10 µL of Plus reagent; DNA was mixed well prior to addition of Plus reagent; (ii) tube B, 250 µL of Opti-MEM medium and 16 µL of Plus reagent.

The mixtures were incubated for 10 minutes at room temperature and then the contents of Tube B were added to Tube A. The DNA:Lipofectamine complexes were allowed to form for 30 minutes. Opti-MEM was added for a final volume of 500 µL. The mixture was then added dropwise to HEK FRT cells previously washed twice with Opti-MEM. After 4 hours, cells were washed three times with Opti-MEM to remove DNA:Lipofectamine complexes and complete DMEM medium was added. BAC transfection efficiency was assessed by X-gal staining (section 2.4.2.1) after 72 hours.

2.4.2 Generation of stable clones by site-specific integration

Stable clonal cell lines were obtained by transfecting HEK FRT cells with either pBAC-*FXN-MS2-Luc* or pBAC-*FXN-GAA-MS2-Luc* together with Flp recombinase-encoding plasmid pOG44 (800 ng, Life Technologies), generating the lines *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc*, respectively. Cells used as a positive control were co-transfected with pH-FRT-Hy (730 ng) and pOG44 (800 ng) and cells used as a negative control were transfected with pH-FRT-Hy only. After 48 hours, each 6 cm dish was split into two 10 cm dishes. Stable selection was performed with 100 µg/ml hygromycin (Life Technologies). Cells were cultured for two weeks, changing the medium regularly to allow healthy clone formation. When identifiable, individual clones were picked into 96-well plates and allowed to grow up to 12-well plates for freezing and screening. Vector integration at the FRT docking site of HEK FRT cells was assessed by PCR, as described in section 2.3.4.1.

2.4.2.1 X-gal staining

X-gal (5-bromo-4-chloro-3-indolyl-D-galactopyranoside) staining was used to determine transfection efficiency of pBAC-*FXN-MS2-Luc* and pBAC-*FXN-GAA-MS2-Luc*. Cells were washed once in PBS and incubated in fixative solution (5% formaldehyde, 0.2% glutaraldehyde in PBS) for 10 minutes. Cells were washed twice and incubated with 1 mg/mL X-gal (Promega) in staining solution (5 mM potassium ferricyanide [$K_3Fe(CN)_6$], 5 mM potassium ferrocyanide [$K_4Fe(CN)_6 \cdot 3H_20$], 2 mM $MgCl_2$ in PBS) at 37 °C overnight.

2.5 Epigenetic analysis

2.5.1 Bisulphite sequencing

Genomic DNA of the *FXN-Luc* and *FXN-GAA-Luc* cell lines was isolated using the Illustra Tissue and Cells GenomicPrep Mini Spin Kit (GE Healthcare). Two micrograms of DNA was used in the bisulfite conversion reaction using the EpiTect Bisulfite kit (Qiagen), according to the manufacturer's instructions. Nested PCR was performed on bisulfite-converted DNA using the following primers (Al-Mahdawi *et al*, 2008): SL1F1 and SL1R1 (first round PCR) and SL1F2 and SL1R2 (second round PCR) for the *FXN* promoter region; F1G and R1G (first round PCR) and F2G and R2G (second round PCR) for the upstream region of the GAA repeats in intron 1 of the *FXN* gene; NH1F and SLGR2 (first round PCR) and NH2F and SLGR1 (second round PCR) for the downstream region of the GAA repeats in intron 1 of the *FXN* gene. In the first round of the nested PCR, 3 µl of bisulfite-converted DNA were denatured at 95ºC for 15 minutes, followed by 35 cycles of denaturation at 94ºC for 30 seconds, primer annealing at 50ºC for 30 seconds and extension at 72ºC for 1 minute and final extension at 72ºC for 10 minutes. In the second round of nested PCR, 5 µl of the first round were used with the same termocycling conditions. Each reaction was carried out in 25 µL containing 1x PCR Buffer, 125 µM of each dNTP (Life Tecnologies), 0.5 µM of each primer, 1x Q Solution (Qiagen) and 1 U of HotStart Taq DNA Polymerase (Qiagen). PCR products were cloned into pGEM-T Easy vector (Promega) and electroporated into *E.coli* DH10B (Life Tecnologies). Recombinant clones were directly identified through blue/white screening and confirmed by *Eco*RI restriction digestion which releases cloned inserts. A total of 10 colonies were sequenced per region and for each cell line using T7 and SP6 primers.

2.5.2 Chromatin Immunoprecipitation

Analysis of histone modifications in the *FXN* promoter and regions flanking GAA repeats in lymphoblastoid (GM15850 and GM15851), *FXN-Luc* and *FXN-GAA-Luc* cell lines was performed as previously described (Carey *et al*, 2009). Lymphoblastoid cell lines were treated with DMSO or C5 for 48 hours before crosslinking. Proteins were cross-linked to DNA by 1% formaldehyde treatment for 4 minutes (GM15850 and GM15851) or 7 minutes (*FXN-Luc* and *FXN-GAA-Luc* cells). Chromatin from lysed cells was sheared by sonication to obtain fragments from 100 to 800 bp using a Bioruptor (Diagenode). Immunoprecipitation experiments were performed using the Immunoprecipitation Kit with

Dynabeads Protein G (Life Technologies), according to the manufacturer's instructions. Chromatin was incubated with one of the following antibodies at 4°C: (i) anti-H3K9ac (07-352), (ii) anti-H4K8ac (07-328), (iii) anti-H3K9me2 (07-441), (v) anti-H3K9me3 (07-442) and (vi) normal rabbit serum (12-370) used as a negative control (all antibodies were purchased from Millipore). Samples were treated with 500 mM NaCl, 0.25 mg/mL RNase A and 0.25 mg/mL proteinase K and incubated at 65°C overnight to reverse crosslinking. After reversing the crosslinking, the amount of *FXN* DNA immunoprecipitated was quantified in triplicate by real-time PCR using the SYBR Green PCR Master Mix (Applied Biosystems) and determined using the ΔΔCt method. The immunoprecipitated DNA was normalised to 10% of input and taking into consideration the background signal. For the analysis of the *FXN* promoter and the regions upstream and downstream of the GAA repeats, primers described by Soragni and colleagues were used (Soragni *et al*, 2008).

2.6 Analysis of *FXN* expression

2.6.1 RT-PCR

Total RNA from HEK FRT, *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* clonal cell lines and from the lymphoblastoid-derived cell lines GM15851, GM14519 and GM16209 was extracted using RNeasy Mini Kit (Qiagen) and treated with RNase-Free DNase (Qiagen). cDNA was synthesized from 1 μg of total RNA using random primers (Life Technologies) and SuperScript III Reverse Transcriptase (Life Technologies) in a total volume of 20 μL as follows: 1.0 μg of RNA, 1.0 μL of random primers (150 ng), 1.0 μL of dNTP (10 mM) and DEPC water up to 13 μL. This mixture was heated to 65 °C for 5 minutes. Next, the following reagents were added: 4 μL of 5x first strand buffer, 1 μL of DTT (0.1M), 1 μL of RNase OUT and 1 μL of Superscript III Reverse Transcriptase (all Life Technologies). The sample was incubated in the thermocycler using the following conditions: 25 °C for 5 minutes, 55 °C for 60 minutes and 70 °C for 15 minutes. RT-PCR amplification was performed as described in section 2.3.4.1 using 5 μL of a 1:10 dilution of cDNA to assess transient expression of *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* transgenes in HEK FRT cells. *GAPDH* expression was used as a positive control.

2.6.2 qRT-PCR

Total RNA from HEK FRT, *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* clonal cell lines and from the lymphoblastoid-derived cell lines GM15851, GM14519 and GM16209 was

extracted using RNeasy Mini Kit (Qiagen) and treated with RNase-free DNase (Qiagen). cDNA was synthesized from 1 µg of total RNA using random primers (Life Technologies) and SuperScript III Reverse Transcriptase (Life Technologies) as described in section 2.6.1. Three independent cDNA samples per cell line were quantified in triplicate by real-time PCR using the SYBR Green PCR Master Mix (Applied Biosystems) in a reaction volume of 25 µL. *FXN-MS2-Luc* mRNA expression was determined using primers qFXN-Luc-F (5′-CGGAAAAGATGCTGGAAGTG-3′) and qFXN-Luc-R (5′-AACCAGGGC GTATCTCTTCA-3′) normalised by *GAPDH* (Forward, 5′-GGTCTCCTCTGACT TCAACA-3′; Reverse, 5′-AGCCAAATTCGTTGTCATAC-3′; RTPrimer DB, ID: 912). *FXN* mRNA expression was detected using primers FXN-F and FXN-R (Herman *et al*, 2006) normalised by *β-actin* (Forward, 5′-AGCGCGGCTACAGCTTCA-3′; Reverse, 5′-CGTAGCACAGCTTCTCCTTAATGTC-3′; RTPrimerDB, ID: 2203).

2.7 Analysis of frataxin protein levels

2.7.1 Luciferase assay

To assay luciferase expression, $1.5x10^6$ cells of *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* clonal cell lines were used. Cells were seeded in 6-well plates, washed in PBS and lysed in Lysis Buffer (25 mM TrisPO_4 pH7.8, 2 mM CDTA, 10% Glycerol and 1% Triton-X 100) for 20 minutes at 4°C. Seventy-five microlitres of lysates were mixed with 100 µL of Luciferase Assay Buffer (15 mM $MgSO_4$, 15 mM KPO_4 pH 7.8, 4 mM ethylene glycol tetraacetic acid [EGTA] pH 7.8, 2 mM adenosine 5′-triphosphate [ATP] and 2mM dithiothreitol [DTT]) and 50 µL of D-Luciferin (0.3 mg/mL). The relative light units of luciferase of each cell line were determined using the Dynex MLX 96 Well Plate Luminometer and were normalised by total protein concentration, determined using bicinchoninic acid solution (BCA, Sigma), according with the manufacturer's instructions.

2.8 Fluorescence *in situ* hybridisation (FISH)

2.8.1 Probes for DNA FISH

Human *FXN* gene was detected in GM15851, GM14519 and GM16209 cells using either G248P87590H8 or G248P89980G8 fosmid clones obtained from BACPAC Resources Center (Children's Hospital Oakland Research Institute). Probes were labelled by nick translation with digoxigenin (Roche). Briefly, 1µg of DNA from each fosmid clone

was treated with 200 ng RNase (Sigma) for 30 minutes at 37°C. After RNA removal, each probe was mixed with 5 µL of nick translation buffer (0.5 M Tris–HCl pH 8.0, 50 mM $MgCl_2$, 0.5 mg/mL BSA), 5 µL of β-mercaptoethanol, 5 µL of dNTP mix (0.5 mM deoxyadenosine triphosphate [dATP], 0.5 mM deoxyguanosine triphosphate [dGTP], 0.5 mM deoxycytidine triphosphate [dCTP]), 1 nmol of digoxigenin, 6 units DNAse I (Roche) and 10 units of DNA polymerase I (Life Technologies) and incubated for 2 hours at 16°C. Unincorporated nucleotides were removed using Illustra G-50 columns (GE Healthcare), according to the manufacturer's instructions. In carrier cells, the expanded allele was distinguished from the wild-type allele using biotinylated $(GAA)_{15}$ or $(TTC)_{15}$ oligos (Biomers). Chromosome 9 was detected using a green XCyting 9 whole chromosome paint for HSA9 (MetaSystems).

2.8.2 2D FISH

DNA FISH was performed on GM14519 carrier cell line as described previously with slight modifications (Brown *et al*, 2008). Cells were fixed in 3:1 methanol–acetic acid and treated with 100 µg/mL RNase for 30 minutes at 37°C, washed in 2x SSC, and dehydrated. Cells were denatured in 70% formamide/2x SSC at 72°C for 2 minutes and dehydrated. Labelled oligos (100 ng) were denatured in FISH hybridization buffer (Kreatech) at 90°C for 8 minutes and immediately allowed to cool on ice. Labelled fosmid clones (100 ng) were precipitated with 3 µg of human Cot-1 DNA (Life Technologies) and 20 µg of salmon sperm DNA (Sigma), denatured in FISH hybridization buffer at 90°C for 8 minutes and were preannealed at 37°C for 8 minutes. Whole chromosome 9 paint (4 µL) was denatured at 75°C for 2 minutes and combined with 8 µL of denatured oligo and fosmid and applied onto the slide. Slides were hybridised overnight at 37°C. After hybridisation, slides were washed in 0.4x SSC/0.3% Igepal for 2 minutes and then 1 minute at room temperature. Slides were blocked in 3% BSA/4x SSC for 30 minutes, biotin was detected with streptavidin Cy3 (Jackson ImmunoResearch Laboratories, 1/200 dilution), and digoxygenin was detected with sheep anti-digoxigenin FITC (Roche, 1/50 dilution) followed by rabbit anti–sheep FITC (Vector Laboratories, 1/100 dilution) or mouse anti-digoxigenin (Jackson ImmunoResearch Laboratories, 1/100 dilution) followed by Alexa Fluor 647 goat anti–mouse IgG (Life Technologies, A-21236, 1/200 dilution) when the green whole chromosome 9 paint was used. Cells were mounted in Vectashield (Vector Laboratories) with 1 µg/mL 4′,6-diamidino-2- phenylindole (DAPI) counterstain.

2.8.3 3D FISH

3D DNA FISH was performed on GM14519 carrier cell line as described previously with slight modifications (Brown *et al*, 2008). Cells were washed in PBS and allowed to settle on poly-L-lysine-treated coverslips for 5 minutes. Cells were fixed in 4% paraformaldehyde (PFA)/25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (Hepes) for 15 minutes and permeabilised in 0.5% Triton X-100 (Calbiochem) in PBS for 10 minutes. RNA was removed with 100 µg/mL RNase in 2x SSC for 1 h at 37°C. Cells were denatured in 3 N HCl for 20 minutes, and neutralized in ice-cold PBS. Probes were prepared as in the previous section, and slides were hybridised overnight at 37°C. Cells were washed two times in 2x SSC at 37°C, once in 1x SSC at room temperature for 30 minutes, and blocked in 3% BSA/4x SSC for 1 hour. Probes were detected as in the previous section. Cells were washes between layers in 4x SSC with 0.05% Tween20. Coverslips were mounted in Vectashield with 1 µg/mL DAPI counterstain.

2.8.4 Immuno-FISH

Immuno-FISH was performed on *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells, and on the lymphoblastoid cell lines GM14519, GM15851 and GM16209 as described previously with slight modifications (Brown *et al*, 2008). Adherent cells were seeded on poly-L-lysine-treated coverslips and allowed to grow overnight. Suspension cells were washed in PBS and allowed to settle on poly-L-lysine-treated coverslips for 5 minutes. Cells were fixed in 4% PFA/25 mM Hepes for 15 minutes and permeabilised in 0.5% Triton X-100 for 10 minutes. Non-specific sites were blocked with 10% FBS. Antibodies were prepared in blocking solution. The NL was detected with goat anti-lamin B1 (Santa Cruz, sc-6217, 1/200 dilution). Nuclear speckles were detected with mouse anti-SC35 (Sigma, S4045, 1/1000 dilution). The nucleolus was detected with mouse anti-nucleolin (Abcam, ab13541, 1/25 dilution). Secondary antibodies used were Alexa Fluor 647 goat anti–mouse IgG (Life Technologies, A-21236, 1/1000 dilution) and Alexa Fluor 647 rabbit anti-goat IgG (Life Technologies, A-21446, 1/500 dilution). After protein detection, slides were post-fixed in 4% PFA/25mM Hepes at room temperature for 10 minutes. Cells were treated with 100 µg/mL RNase for 1 h at 37°C, denatured in 3 N HCl for 20 minutes, and neutralized in ice-cold PBS. Probes were prepared as in section 2.8.2, and slides were hybridised overnight at 37°C. Cells were washed twice in 2x SSC at 37°C, once in 1x SSC at room tempearture, and blocked in 3% BSA/4x SSC. Probes were detected as in section 2.8.3.

2.8.5 RNA FISH

RNA FISH was performed on HEK FRT and *FXN-MS2-Luc* cell lines as previously described (Chartrand *et al*, 2000) with minor modifications. Briefly, cells were allowed to grow on poly-L-lysine-treated coverslips overnight, washed in PBS and fixed in 4% formaldehyde (Electron Microscopy Science) for 10 minutes at room temperature. Cells were permeabilised in 0.5% Triton X-100 for 6 minutes and washed in 50% formamide/2x SSC for 5 minutes. Cells were hybridised overnight at 37°C with 10 ng Cy3-MS2 DNA probe (Hocine *et al*, 2013) in 40 µL hybridisation mix containing 2x SSC, 3% BSA, 10% dextran sulphate (Sigma), 2 mM vanadyl-ribonucleoside complex (New England Biolabs), 40 µg tRNA (Roche), and 50% formamide. Cells were washed once in 50% formamide/2x SSC at 37°C for 10 minutes, once in 2x SSC for 10 minutes at 37°C and once in 2x SSC for 5 minutes at room temperature. Cells were washed briefly in PBS and mounted in Vectashield with 1 µg/mL DAPI counterstain.

RNA FISH on GM15851 and GM16209 cell lines was performed as described previously (Raj & Tyagi, 2010) with minor modifications. Custom Stellaris FISH probes TSS (20 oligos), UpGAA (41 oligos) and DownGAA (41 oligos) were designed to hybridise with the regions TSS-exon1, upstream and downstream of the GAA repeats, respectively (Table 2.1), according to the manufacturer's instructions. Cells were washed in PBS and allowed to settle on poly-L-lysine-treated coverslips for 5 minutes. Cells were fixed in 3.7% formaldehyde/1X PBS for 10 minutes, permeabilised in 70% ethanol at 4°C for 1 hour and washed in 10% formamide/2x SSC at room temperature for 5 minutes. Hybridisation was performed overnight at 42°C with probes TSS (50 nM), UpGAA (50 nM) and DownGAA (25 nM) in 100 µL hybridisation buffer containing 100 mg/mL dextran sulphate, 10% formamide in 2x SSC. Cells were washed in 10% formamide/2x SSC at 42°C for 10 minutes. Cells were washed briefly in PBS and mounted in Vectashield with 1 µg/mL DAPI counterstain.

Table 2.1 Custom Stellaris FISH probes TSS, UpGAA and DownGAA used in RNA FISH experiments.

Probe #	TSS	UpGAA	DownGAA
1	acccctgacccaagggagac	ctgagaaggagcggggtgag	tttttgagacggagtcttgc
2	aaagcacggagtgcaaccag	ccagagaagggagttgcaag	gctgggaaatccattttatt
3	aaatggagagcctgctttgt	cagtgtaaatgcaaccggga	acttgcctatttttccagag
4	gcactattcgtgcatttaac	cttttccttcggaaagcaga	gatctaaggaccatcatggc
5	caggaagaacttcccagctt	tcgcaccgcaggacaaaatg	aaatgtctgctttcctagag
6	gagcagctagaggttagacc	tcgcagagaagtgacaagca	aatacgtggcagctcagata
7	caggcactcttctgtggg	actagcagataccatctatt	acacaactctgctgacaacc
8	ccctggtggccactggcc	cagacacacacattatgtgt	cctgggcaacaagagggaaa
9	cgctgggtgctgcggcga	acacacacgctatatacaga	tgcagtgagctgagactgag
10	cgcccgctccgccctcca	tgcgcgcaaacacacacaca	cttgaacccaggaggcagag
11	catgctgctccgggtctg	aaatattaggtgtgcgcgcg	ggaggctgaggcaggagaat
12	cgcggcgcccgagagtcc	ccaggaaagcatttcgttca	aattagccgataatcccagc
13	gccaggaggccggctact	gctctgaaagtttcacctcg	gcctgaccaacatggagaaa
14	ctgggccgggctgggtga	ctgctgctctagctattctg	tgaggttgggagttcgagac
15	cccgggtgagggtctggg	gtcgggtcagtttccaaaag	taattccagcactttgggag
16	aactctgccggccgcggg	gtggggcagaatctggaata	aaaaactctgctgatggcca
17	gcggccgcagagtggggc	caaggtcacacagctctgcg	cctcttgtcattctcctaaa
18	tgtcggtgcgcaggccac	ctcagagaggttaggggaat	ccaccaggtaccatctaaaa
19	ggcgtgcaggtcgcatcg	ccctacagaaaacaaagcca	aagttgggtgtgtttgtcag
20	gatacttactgcgcggcg	caaaatgggcgtcaccttta	tatccagaaacccagtgtcc
21		tttaatcctcacaccaggtc	gtgacagaggttgctaacca
22		gactttatctatgttattcc	tagaaaaaaagaggcccagc
23		cccaaaggaagaaaggggaa	gtgcaagggaactatggaac
24		ggtcagatatttctttgtac	caactcctcaaatacccaag
25		ctttcaagccgtggcgtaac	aattaaacaggatggtggcc
26		acagccattctttgggtttc	aggacagtcagggctttaaa
27		ttgaggaatcttcctcatcc	tgaattttggagaccagggg
28		cattaaataccatgtcctcc	gccttcaagtgaactcagtt
29		ttccttggcatcttcaagac	gtggagaaaagggtggggaa
30		ctcgtgaaacaccctctacc	ctgctttagaagtagatgca
31		ttcatctcccctaatacatg	ccattctgtttctgttgaac
32		aaatatggcttggacgtggc	tgtagaattatgtgtgtggc
33		gtacaaactccggagagcaa	cctggtctcattactgattt
34		gtgtgggaagtttaagccta	caaagtcctgggattacagt
35		cacaatgtgggccaaataac	cttgaactcctgatctcaga
36		ccaatcccaaagtttcttca	agagatggggtttcaccatg
37		cctaacttttaagcactggc	ccatgctgagctagttttta
38		gccaggaaatccattttcta	cgagtagctgggactacagg
39		aagtgctgagattatgggca	aagtgattctcatgcctcag
40		caggttagtcttgaactccg	aaatcttggctcactgcaac
41		tagatactgggtttcaccat	tttgtatgtcacagaatccc

2.9 Fluorescence Recovery After Photobleaching (FRAP)

FRAP experiments were performed using the UltraView Spinning Disk system (PerkinElmer) mounted on an IX81 microscope (Olympus) equipped with a 100× 1.3 numerical aperture (NA) objective, a FRAP unit and an electron multiplying charge-coupled device (EMCCD) camera (Hamamatsu Photonics) and driven by Volocity software (PerkinElmer). HEK FRT and *FXN-MS2-Luc* cell lines were seeded on collagen-coated glass-bottom dishes (MatTek) in complete DMEM medium (section 2.1.1) and transiently transfected with NES-YFP-MS2-NLS (a kind gift from Y. Shav-Tal) (Ben-Ari *et al*, 2010). For nuclear detection, NucBlue Live reagent was added to the cells 20 minutes before imaging. Cells were maintained at 37 °C in a 5% CO_2 humidified atmosphere using an on-scope incubator and an objective heater (Tokai). Fluorescent spots inside the nucleus indicating active *FXN-MS2-Luc* transcription sites were detected with the YFP channel (514 nm laser). The nucleus was detected with the DAPI channel (405 nm laser). z-stacks were acquired at 0.25 µm intervals to cover the entire nuclear volume of each cell. Before bleaching, three z-stacks of each cell were taken. The active transcription site was bleached for 200 ms with the 514 nm laser power set to 100%. Post-bleach recovery z-stacks images were acquired with an exposure time of 50 ms every 20 seconds for 20 minutes.

2.10 Imaging acquisition and analysis

Images of fixed cells were obtained using the DV Elite system based on an Olympus IX71 fully motorised widefield deconvolution inverted microscope with a 100x objective 1.40 NA fitted with a CoolSNAP HQ^2 cooled charge-coupled device (CCD) camera (Photometrics) driven by SoftWoRx 5.0 software (Applied Precision). Several cell positions were chosen randomly and recorded using the motorized stage. z-stacks were acquired at 0.2 µm intervals to cover the entire volume of each cell. Distance measurements between FISH signals and the nuclear periphery were calculated with NEMO software (Iannuccelli *et al*, 2010). Briefly, a region of interest (ROI) was defined around each nucleus using an image stack from the DAPI channel. The software applied several segmentation filters (Median 3D, TopHat 3D, mathematical morphology 3D) on raw image stacks for object detection. Objects were detected by computing seeds as the brightest maximum local pixels above a global threshold, and a local threshold was then computed to aggregate neighbouring pixels to the seed. For each cell, the distance of each allele to the nuclear periphery was normalised by the local ray between the nuclear centre and periphery passing through the FISH signal. For relative positional analysis, image

stacks were deconvolved using the SoftWoRx 5.0 software. Association of FISH signals with subnuclear domains or chromosome territories was analysed using ImageJ (http://imagej.nih. gov/ij/): image stacks were analysed manually and association was determined by eye when *FXN* FISH signals were touching or superimposing the signal from a specific subnuclear domain at a given z-section. Images are maximum projections of z-stacks covering the signals. Contrast-stretch and gamma adjustments were made using Adobe Photoshop CS5 only for display in Figures 5.4B, 5.5D and 5.6A.

RNA MS2 FISH signals were detected as described previously (Raj *et al*, 2008) using StarSearch software (Levesque & Raj, 2013). Briefly, the program applied Laplacian of Gaussian filters in three-dimensions, taking into consideration the size of the fluorescent spots to remove the non-uniform background while enhancing particles. After filtering, the program calculated the number of spots detected for all possible thresholds. After plotting this data, the graph showed a plateau region corresponding to a broad range of thresholds over which the spot count did not vary significantly. Spot counts calculated with these thresholds matched the number of spots identified by eye. In HEK FRT cells and cells from the *FXN-MS2-Luc* cell lines that did not express the transgene, the plateau region was not observed. The DIC channel was used to determine the cell edges.

In RNA FISH experiments on GM15851 and GM16209 cell lines several cell positions were chosen randomly. z-stacks were acquired at 0.2 µm intervals to cover the entire volume of each nucleus. Images were analysed using ImageJ. For the analysis of nascent RNA FISH signals using the UpGAA probe only, a threshold was defined for both cell lines to eliminate non-specific background. RNA fluorescent spots were identified either at the nuclear rim or interior using the DAPI channel and their intensity measured using the 3D Object Counter plugin (Bolte & Cordelières, 2006). Identification of RNA FISH signals using probes TSS, UpGAA and DownGAA was performed as described above. In addition to threshold determination, to distinguish true signals from background, the downstream signals were scored when co-localised with the upstream signals. Because the TSS probe hybridises to exon 1, it could potentially label mature *FXN* mRNA. However, due to its short length spanning, the TSS probe did not generate more than two RNA FISH signals per cell. Therefore, all TSS RNA FISH signals were identified as nascent transcript accumulating at the transcription sites.

For live-cell experiments, HEK FRT and *FXN-MS2-Luc* cell lines were seeded on collagen coated glass-bottom dishes (MatTek) in complete DMEM medium and transiently transfected with NES-YFP-MS2-NLS (Ben-Ari *et al*, 2010). For live-cell imaging of

transcription sites, cells were imaged in four dimensions (three dimensions over time) using the DV Elite system based on an Olympus IX71 microscope with a 100x objective 1.40 NA fitted with an Evolve EMCCD camera (Photometrics) driven by SoftWoRx 5.0 software under physiological conditions inside an incubator at 37°C and 5% CO_2.

FRAP image sequences were analysed with Volocity software (PerkinElmer). Tracking of fluorescence spots was performed using the manual tracking function on Volocity at each time point. Relative intensity of each time point was calculated using

$$I_{\mathrm{rel}} = (T_0 - B_0) \times \frac{I_t - B_t}{T_t - B_t} \times (I_0 - B_0)$$

where T_0 and $T_{(t)}$ are the average intensity of an arbitrary area in the nucleus before and after bleaching, respectively, and I_0 and $I_{(t)}$ are the average intensity of the ROI before and after bleaching, respectively. The background was subtracted from all measured values. The formula accounted for fluorescence loss due to bleaching monitoring, and finally fluorescence intensity was normalised to 1 (Dundr & Misteli, 2003). Data from five cells per lines were collected in three independent experiments, and the averaged FRAP measurements were fitted with exponential curves. The diffusion of MCP-YFP protein during FRAP analysis was disregarded as the diffusion rate of free MCP-YFP is very rapid, whereas the bound MCP-YFP is associated with high affinity to the mRNA and does not detach, diffuse and bind again at the transcription site (Yunger *et al*, 2010).

CHAPTER 3

Epigenetic silencing induced by a GAA repeat expansion in a human cell model of FRDA

The results featured in this Chapter were published in the paper (Annex 1):

Lufino MMP, **Silva AM,** Nemeth AH, Alegre-Albarrategui J, Russell AJ, Wade-Martins R (2013) A GAA repeat expansion reporter model of Friedreich's Ataxia recapitulates the genomic context and allows rapid screening of therapeutic compounds. *Hum. Mol. Genet.* 22(25): 5173-87.

Author contributions:

M.M.L., **A.M.S** and R.W.-M. conceived and designed the experiments; M.M.L and **A.M.S** performed the experiments and analysed the data; A.H.N, J.A.A and A.J.R. provided reagents; M.M.L. and **A.M.S** wrote the manuscript; M.M.L., **A.M.S**, A.H.N, A.J.R. and R.W.-M. revised the manuscript.

3.1 Introduction

The generation of FRDA models has been essential to uncover (i) the molecular mechanisms underlying the GAA expansion-mediated gene silencing and instability, (ii) the pathophysiology of the disease, (iii) the function of the frataxin protein, (iv) new therapeutic targets and (v) potential therapeutic compounds (reviewed in Martelli *et al*, 2012; Perdomini *et al*, 2013; Yandim *et al*, 2013; Sandi *et al*, 2014a).

An ideal FRDA model should have a GAA repeat expansion within the *FXN* genomic context, leading to impairment of *FXN* transcription. Cellular models should recapitulate the tissue-specific disease phenotype and GAA repeat instability. Additionally, animal models should recapitulate FRDA biochemical and cellular defects and disease progression (Perdomini *et al*, 2013). So far, none of the available FRDA models meets all these criteria, so the use of a specific model depends on the scientific question being answered. Here, I will focus on the epigenetic findings from available FRDA cellular and animal models used to understand the pathological molecular mechanism caused by expanded GAA repeats and to test compounds able to upregulate *FXN* expression.

3.1.1 FRDA cell models

Several studies have been conducted using patient-derived lymphoblastoid cells, primary fibroblasts or peripheral lymphocytes. Although these cells recapitulate some molecular features of FRDA, including the GAA-mediated *FXN* heterochromatinization and decreased levels of *FXN* mRNA, they do not develop spontaneous phenotypic characteristics of the disease (Rotig *et al*, 1997; Sturm *et al*, 2005).

FRDA-iPSC lines have been developed in order to generate patient-specific and disease-relevant cell lineages, including neurons and cardiomyocytes. FRDA iPSCs recapitulate some of the molecular and genetic aspects of FRDA. The pathological GAA repeat expansion in iPSCs decrease the levels of *FXN* mRNA and frataxin protein, but do not lead to a deficit of the activity of ISC proteins (Perdomini *et al*, 2013).

3.1.2 FRDA mouse models

GAA-based mouse models were obtained using either a knock-in approach, based on homologous recombination, or a YAC transgenic approach.

Homozygous 230 GAA repeat expansion knock-in mice (KIKI) show a 25% decrease in frataxin expression, whereas compound heterozygous knock-in–knockout mice

(KIKO) express 25-35% of wild-type frataxin levels (reviewed in Perdomini *et al*, 2013). Although only presenting a mild FRDA phenotype, these models have been useful to investigate epigenetic-based *FXN* upregulating compounds, such as HDAC inhibitors (Rai *et al*, 2008, 2010).

The human *FXN* YAC transgenic mouse models express human *FXN* in a mouse *FXN*-null background: Y47R, contains $(GAA)_9$ repeats, and YG8R and YG22R, which initially contained 90-190 and 190 GAA repeats, respectively, now contain 120-220 units and 170-260 GAA repeats, respectively (Pook *et al*, 2001; Al-Mahdawi *et al*, 2004, 2008; Anjomani Virmouni *et al*, 2014). Both YG8R and YG22R mice express decreased levels of human *FXN* mRNA and frataxin protein in comparison with wild-type or Y47R control mice (Al-Mahdawi *et al*, 2004, 2008; Anjomani Virmouni *et al*, 2014). Furthermore, both YG8R and YG22R mice exhibit a progressive FRDA-like molecular disease phenotype, including intergenerational and somatic instability of the GAA repeat expansion (Al-Mahdawi *et al*, 2004; Clark *et al*, 2004) and mild progressive behavioural motor coordination deficits when compared with wild-type or Y47R controls (Al-Mahdawi *et al*, 2008; Anjomani Virmouni *et al*, 2014). Recently, a new line designated YG8sR was generated from YG8R and contains a single copy of the *FXN* transgene with ~200 pure GAA repeats (Anjomani Virmouni *et al*, 2015).

There are also several condicitional *FXN* knockout models which exhibit severe early-onset FRDA-like pathologies, being extremely useful for assessing potential FRDA drug therapies acting on the downstream effects of frataxin deficiency (Puccio *et al*, 2001; Martelli *et al*, 2012).

3.1.3 Epigenetic changes at the expanded *FXN* allele

The GAA repeat expansion has been shown to induce epigenetic changes at the *FXN* locus, including DNA methylation and hydroxymethylation, histone deacetylation, and histone methylation, which may lead to reduced *FXN* mRNA levels in FRDA patient cells, tissues, cellular and mouse models (Table 3.1).

Studies have revealed increased DNA methylation of specific CpG sites in the upstream region of GAA repeat expansions in FRDA patient lymphoblastoid cell lines, FRDA patient tissues and YG8 and YG22 *FXN* transgenic mice (Greene *et al*, 2007; Al-Mahdawi *et al*, 2008). In particular, three (CpG sites 3, 6 and 13) out of 15 CpG sites were found to be usually unmethylated in healthy cells, but highly methylated in FRDA cells. One of these residues is located within an E-box whose deletion caused a significant drop

in *FXN* promoter activity in reporter assays (Greene *et al*, 2007). Furthermore, Al-Mahdawi and colleagues showed that the increased level of DNA methylation at CpG site 12 predominantly comprises 5hmC rather than 5mC (Al-Mahdawi *et al*, 2013). The degree of DNA methylation in the upstream GAA repeat region has been shown to correlate with the length of the GAA repeats and inversely correlate with *FXN* expression levels and age of disease onset (Castaldo *et al*, 2008; Evans-Galea *et al*, 2012).

Table 3.1 Epigenetic changes in FRDA patient cells, tissue, cellular and mouse models (Yandim *et al*, 2013; Sandi *et al*, 2014a)**.** FRDA results were compared to levels obtained on cells or tissues derived from healthy individuals. A relatively high level compared to healthy is marked with '↑', low level with '↓'. For YG8R and YG22R *FXN* YAC transgenic mice with expanded GAA repeats, levels were compared to Y47 transgenic mice containing only have 9 GAA repeats. For HEK293T cells, results obtained from engineered 560 GAA repeats into a *GFP* intron were compared to a control construct without GAA repeats.

Chromatin change	Location	FRDA tissue, cell, or animalmodel	References
5mC ↑	GAA upstream	FRDA patient tissues, primary lymphocytes, lymphoblasts, *FXN* YAC transgenic mice	Greene *et al*, 2007; Al-Mahdawi *et al*, 2008; Evans-Galea *et al*, 2012.
5hmC ↑	GAA upstream	FRDA patient tissues	Al-Mahdawi *et al*, 2013.
H3K4me2/3 ↓	*FXN* promoter/exon 1	Lymphoblasts	Punga & Bühler, 2010; Kim *et al*, 2011; Kumari *et al*, 2011.
	GAA upstream	Lymphoblasts	Kim *et al*, 2011; Kumari *et al*, 2011.
	GAA downstream	Lymphoblasts	Kim *et al*, 2011; Kumari *et al*, 2011.
H3K9me2/3 ↑	*FXN* 5' UTR/promoter	Primary fibroblast, lymphoblasts	De Biase *et al*, 2009; Kim *et al*, 2011.
	GAA upstream	Lymphoblasts, *FXN* YAC transgenic mice, KIKI mice	Herman *et al*, 2006; Al-Mahdawi *et al*, 2008; Rai *et al*, 2008; Punga & Bühler, 2010; Kim *et al*, 2011; Kumari *et al*, 2011; Chan *et al*, 2013.
	GAA downstream	FRDA patient tissues, *FXN* YAC transgenic mice, lymphoblasts	Herman *et al*, 2006; Al-Mahdawi *et al*, 2008; Punga & Bühler, 2010; Kim *et al*, 2011; Kumari *et al*, 2011; Chan *et al*, 2013.
H3K27me3 ↑	*FXN* 5' UTR/promoter	Primary fibroblasts, lymphoblasts	De Biase *et al*, 2009; Kim *et al*, 2011.
	GAA upstream	Lymphoblasts	Kim *et al*, 2011; Chan *et al*, 2013.
	GAA downstream	Lymphoblasts	De Biase *et al*, 2009; Kim *et al*, 2011.
H3K36me3 ↓	GAA upstream	Lymphoblasts	Punga & Bühler, 2010; Kim *et al*, 2011; Kumari *et al*, 2011.
	GAA downstream	Lymphoblasts	Punga & Bühler, 2010; Kim *et al*, 2011; Kumari *et al*, 2011.
H3K79me2 ↓	GAA upstream	Lymphoblasts	Kim *et al*, 2011.
	GAA downstream	Lymphoblasts	Kim *et al*, 2011.

H4K20me3 ↑	GAA upstream	Lymphoblasts	Kim *et al*, 2011.
	GAA downstream	Lymphoblasts	Kim *et al*, 2011.
H3K9ac ↓	*FXN* promoter	FRDA patient brain tissue, lymphoblasts	Al-Mahdawi *et al*, 2008; Kumari *et al*, 2011.
	GAA upstream	FRDA patient tissues, lymphoblasts, HEK GFP-GAA_{560}, *FXN* YAC transgenic mice, KIKI mice	Herman *et al*, 2006; Al-Mahdawi *et al*, 2008; Soragni *et al*, 2008; Kumari *et al*, 2011.
	GAA downstream	FRDA patient tissues, lymphoblasts, HEK GFP-GAA_{560}, *FXN* YAC transgenic mice	Herman *et al*, 2006; Al-Mahdawi *et al*, 2008; Soragni *et al*, 2008; Kumari *et al*, 2011.
H3K14ac ↓	*FXN* promoter	FRDA patient tissues	Al-Mahdawi *et al*, 2008.
	GAA upstream	Lymphoblasts, HEK GFP-GAA_{560}, KIKI mice	Rai *et al*, 2008; Soragni *et al*, 2008; Kumari *et al*, 2011.
	GAA downstream	Lymphoblasts, HEK GFP-GAA_{560}	Soragni *et al*, 2008; Kumari *et al*, 2011.
H4K5ac ↓	*FXN* promoter	Lymphoblasts	Kumari *et al*, 2011.
	GAA upstream	Lymphoblasts, HEK GFP-GAA_{560}, KIKI mice	Herman *et al*, 2006; Rai *et al*, 2008; Soragni *et al*, 2008; Kumari *et al*, 2011.
	GAA downstream	FRDA patient tissues, HEK GFP-GAA_{560}, *FXN* YAC transgenic mice, lymphocytes	Herman *et al*, 2006; Rai *et al*, 2008; Soragni *et al*, 2008; Kumari *et al*, 2011.
H4K8ac ↓	GAA upstream	Lymphoblasts, FRDA patient tissues, HEK GFP-GAA_{560}, KIKI mice	Herman *et al*, 2006; Al-Mahdawi *et al*, 2008; Rai *et al*, 2008; Soragni *et al*, 2008; Kumari *et al*, 2011.
	GAA downstream	Lymphoblasts, FRDA patient tissues, HEK GFP-GAA_{560}	Herman *et al*, 2006; Soragni *et al*, 2008; Kumari *et al*, 2011.
H4K12ac ↓	*FXN* promoter	Lymphoblasts	Herman *et al*, 2006.
	GAA upstream	Lymphoblasts, FRDA patient tissues, *FXN* YAC transgenic mice	Herman *et al*, 2006; Al-Mahdawi *et al*, 2008.
	GAA downstream	Lymphoblasts, FRDA patient tissues, *FXN* YAC transgenic mice	Herman *et al*, 2006; Al-Mahdawi *et al*, 2008.
H4K16ac ↓	*FXN* promoter	Lymphoblasts	Rai *et al*, 2008; Kumari *et al*, 2011.
	GAA upstream	Lymphoblasts, FRDA patient tissues, *FXN* YAC transgenic mice, KIKI mice	Herman *et al*, 2006; Al-Mahdawi *et al*, 2008; Rai *et al*, 2008; Kumari *et al*, 2011.
	GAA downstream	Lymphoblasts, FRDA patient tissues, *FXN* YAC transgenic mice, KIKI mice	Herman *et al*, 2006; Al-Mahdawi *et al*, 2008; Rai *et al*, 2008; Kumari *et al*, 2011.

Hypoacetylation of several lysine residues of histones H3 and H4 (e.g. H3K14ac, H4K5ac, H4K12ac) and di- and trimethylation of H3K9 were first identified at the regions flanking the GAA repeat expansion in FRDA patient-derived lymphoblastoid cells (Herman *et al*, 2006). These epigenetic markers, usually associated with silenced heterochromatin, have since been confirmed in various models, including patient lymphoblast and fibroblast cell lines, iPSCs derived from FRDA patient cells, iPSC-derived neurons, patient autopsy tissues, YG8 and YG22 *FXN* YAC transgenic mice, KIKI mice and cellular models containing integrated transgenes with long GAA repeats (Greene *et al*, 2007; Al-Mahdawi *et al*, 2008; Rai *et al*, 2008; Soragni *et al*, 2008; De Biase *et al*, 2009; Ku *et al*, 2010; Punga & Bühler, 2010; Kim *et al*, 2011; Kumari *et al*, 2011; Sandi *et al*, 2011; Chan *et al*, 2013). Whether GAA-mediated heterochromatin is also present at the *FXN* promotor is still a matter of debate, but a few studies indicate that the *FXN* promoter region of FRDA patient cells show reduced levels of certain marks of active chromatin, as well as reduced occupancy of CTCF and increased levels of FAST1 at the 5' UTR (De Biase *et al*, 2009; Kumari *et al*, 2011). Increased expression of FAST1 in FRDA mouse fibroblasts (Sandi *et al*, 2014b) and reduced CTCF binding at the 5' UTR region of the *FXN* gene in FRDA cerebellum tissue (Al-Mahdawi *et al*, 2013) have also been reported.

Given the GAA-mediated heterochromatin at expanded *FXN* loci, HDAC inhibitors emerged as potential candidates to prevent deacetylation of histones, and thus restore euchromatin and increase *FXN* transcription. A first screening of commercially available HDAC inhibitors was performed on healthy and FRDA patient-derived lymphoblastoid cells, identifying compound 4b which was able to increase both *FXN* mRNA and protein levels in patient cells (Herman *et al*, 2006). Subsequently, three derivatives (compounds 106, 136 and 109) were tested on KIKI mice (Rai *et al*, 2010), *FXN* YAC transgenic mice (Sandi *et al*, 2011) and neuronal cells derived from FRDA iPSCs (Soragni *et al*, 2014). Compound 109 yielded the most promising results. Similarly, the sirtuin protein deacetylase inhibitor nicotinamide (vitamin B3) also increased *FXN* mRNA levels in FRDA cells and *FXN* YAC transgenic mice (Chan *et al*, 2013).

Cellular models used for high-throughput screening of compounds capable of *FXN* upregulation usually contain engineered reporter *FXN* constructs. A stable cell line containing a partial sequence of *FXN* intron 1 with 145 GAA repeats fused to a *GFP* reporter was used for screening of compounds that could increase transcription through the GAA repeat region (Grant *et al*, 2006). Another cell line containing a *GFP* reporter minigene with 560 GAA repeats inserted in an intron recapitulated the molecular hallmarks

of the expanded *FXN* locus, such as reduced mRNA and protein levels, patterns of chromatin modifications around the repeats, and repeat instability (Soragni *et al*, 2008). HDAC inhibitors were able to increase expression of the reporter gene. Additionally, a stable human cell line containing a *FXN-EGFP* fusion construct with the entire normal human *FXN* gene in a BAC clone was generated to screen a chemical library of FDA-approved compounds and natural extracts (Li *et al*, 2013).

3.1.4 Aims of this Chapter

Expanded GAA repeats lead to heterochromatin-mediated dysfunction of the *FXN* gene. In order to better understand the underlying pathogenesis and ultimately to design effective therapies for FRDA, it is important to generate cellular models that recapitulate the repressive hallmarks of the disease while providing efficient ways to quantify *FXN* expression. Here, I describe the characterisation and epigenetic analysis of the first model that allows a direct comparison of the effect of normal or expanded GAA repeats on *FXN* expression within the genomic context of the *FXN* locus.

3.2 Results

I performed all experiments and analysed the results featured in this section, unless stated that an experiment had been previously perfomed in the laboratory.

3.2.1 Characterisation of the *FXN-Luc* and *FXN-GAA-Luc* cell lines

The *FXN-GAA-Luc* and *FXN-Luc* human cell lines were previously generated in the laboratory to allow a direct comparison of the effect of normal and expanded GAA repeats on *FXN* gene expression. These lines contain BAC vectors carrying the 80 kb *FXN* locus, which had been previously modified by homologous recombination to insert either a ~310 GAA repeats expansion (pBAC-*FXN-GAA-Luc*) or 6 GAA repeats (pBAC-*FXN- Luc*), with an insertion of a Luciferase gene in exon 5a, respectively (Fig. 3.1A). Because the native *FXN* promoter and all introns and regulatory sequences required for *FXN* expression are present in the each vector, physiologically relevant *FXN* expression levels can be achieved (Gomez-Sebastian *et al*, 2007). FISH analysis has previously revealed that each cell line possesses the vector integrated at the same location on chromosome 1p.

I assessed vector copy number in *FXN-Luc* and *FXN-GAA-Luc* cells by real-time PCR, using as reference the HEK FRT acceptor cell line (Fig. 3.1B), which carries three endogenous *FXN* loci previously detected by FISH. These stable cell lines carried a single

copy of each vector in the same integration site located on chromosome 1p (Fig. 3.1B and C). Confirmation of the size and purity of the GAA repeat expansion was achieved by PCR, *Mbo*II digestion and sequencing (Fig. 3.1D-F). Sequencing was performed from both sides of the GAA expansion. I confirmed the presence of 213 GAA repeats on the GAA strand and 125 GAA repeats on the TTC strand, thereby achieving an overlap of at least 28 GAA repeats and ensuring that the whole GAA expansion was sequenced. No interruptions were found in the expanded GAA sequence. To further confirm its purity, I performed a *Mbo*II digestion of the PCR product containing the GAA repeats, according to a previously described method (Fig. 3.1E) (Holloway *et al*, 2011). *Mbo*II recognises and cleaves a GAAGA sequence. Complete digestion confirmed the purity of the GAA expansion in the *FXN-GAA-Luc* cell line.

Expansion of GAA repeats causes intergenerational and age-related somatic instability in FRDA patients cells and tissues (De Biase *et al*, 2007a, 2007b) and also in mice tissues (Al-Mahdawi *et al*, 2004; Clark *et al*, 2007). *FXN-GAA-Luc* cells were continuously sub-cultured for 3 months to analyse the degree of the GAA repeat-length variation over time. Instability of the GAA repeat expansion was assessed by PCR (Fig. 3.1G) and small-pool PCR (Fig. 3.1H). PCR analysis of the GAA repeat sequence in the *FXN-GAA-Luc* cell line did not show any difference in size between time points (Fig. 3.1G). I next performed a small-pool PCR analysis to detect uncommon GAA repeat expansion variants that could be present only in a small proportion of cells. However, this too confirmed GAA repeat stability in *FXN-GAA-Luc* cells (Fig. 3.1H). Taken together, these data suggest that the GAA repeat expansion in the *FXN-GAA-Luc* transgene is stable.

The ~310 GAA repeat expansion in *FXN-GAA-Luc* cells was previously shown to reduce *FXN-Luc* mRNA levels by 37% and protein levels by 42%, indicating that the GAA repeat expansion induces *FXN* repression as seen in FRDA.

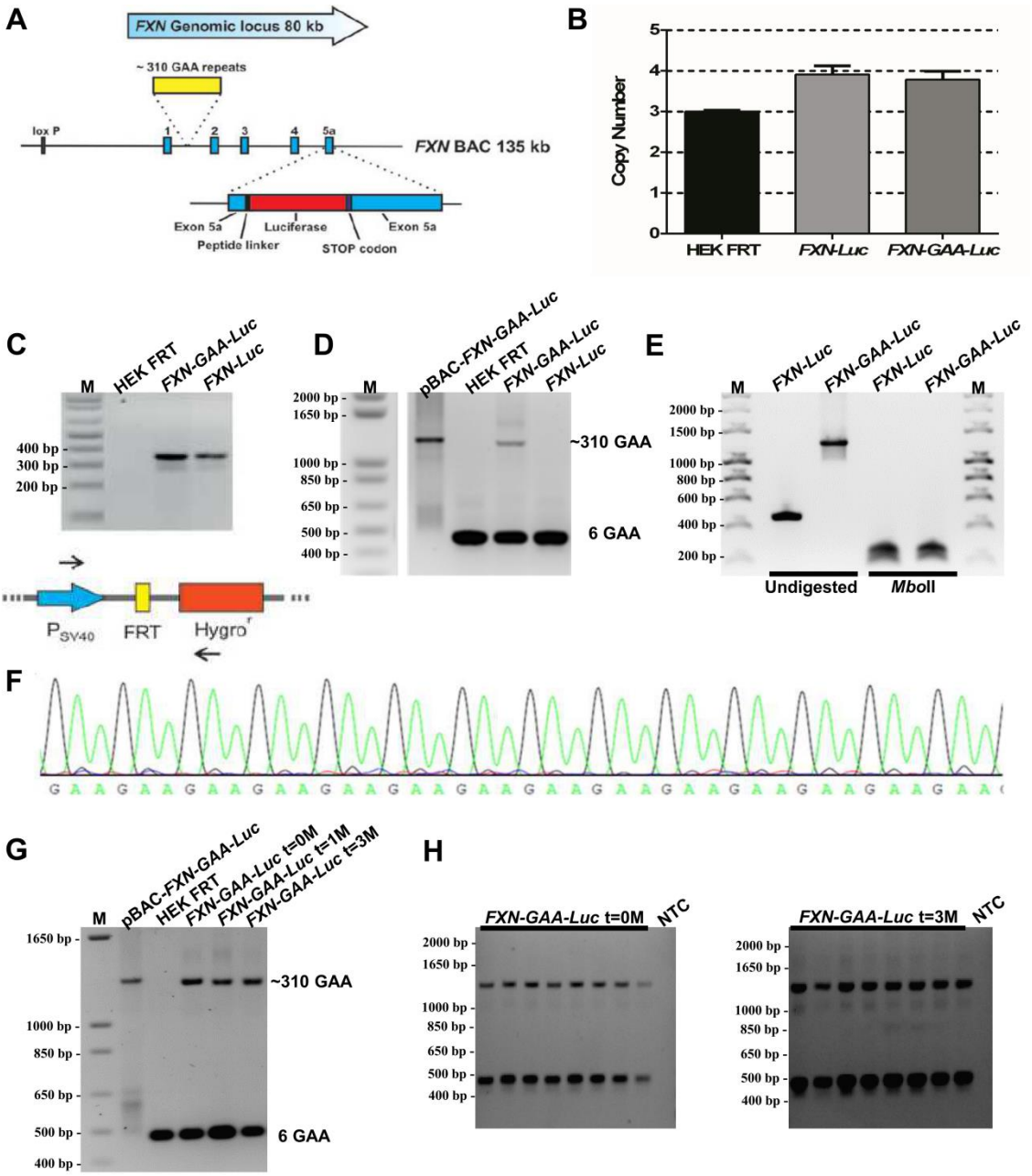

Figure 3.1 Characterisation of the *FXN-GAA-Luc* and *FXN-Luc* cell lines. (**A**) Schematic representation of the pBAC-*FXN-MS2-Luc* and pBAC-*FXN-GAA-MS2-Luc* vectors. Each vector carries either six or ~310 GAA repeats in intron 1, respectively, and a luciferase sequence preceded by a GSGSG peptide linker in exon 5a, immediately before the stop codon. Expression results in a FXN-luciferase fusion protein. (**B**) Single-copy clones were determined by real-time PCR using UpGAA primers normalised by *GAPDH*. Data were relative to HEK FRT, which contains three endogenous copies. Data are mean ± SEM (n = 3). (**C**) Site-specific vector integration in single-copy *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* clones was confirmed by PCR using primers targeting the P_{SV40} promoter and the hygromycin cassette. (**D**) Insertion and sizing of GAA repeats was determined by PCR analysis. (**E**) The purity of the GAA repeat expansion was analysed by *Mbo*II digestion after PCR amplification of the repeats. The resulting fragments of 208 bp and 248 bp contained only the regions flanking the GAA repeats, indicating the GAA repeat sequence was not interrupted. (**F**) The ~310 GAA repeats were sequenced for sizing and purity analysis. A section of the sequencing data is shown. (**G-H**) GAA repeat instability in *FXN-GAA-Luc* cells after continuous sub-culture for 3 months. (**G**) Analysis by PCR did not show any repeat instability. (**H**) For small-pool PCR, eight independent reactions were set up from a starting DNA concentration of 150 pg. The ~310 GAA repeats were stable in *FXN-GAA-Luc* cells. NTC, no template control.

3.2.2 Epigenetic characterisation of *FXN-Luc* and *FXN-GAA-Luc* cell lines

Expanded GAA repeats can induce *FXN* silencing via a heterochromatin-mediated mechanism of repression (Saveliev *et al*, 2003; Herman *et al*, 2006; Chan *et al*, 2013). Epigenetic changes around expanded GAA repeats have been identified, which include increased DNA methylation at specific CpG sites upstream of the GAA repeats (Greene *et al*, 2007; Al-Mahdawi *et al*, 2008; Castaldo *et al*, 2008; Evans-Galea *et al*, 2012) and reduced acetylation of histones H3 and H4 accompanied by increased levels of methylated histones H3K9me2 and H3K9me3 in regions flanking GAA repeats (Herman *et al*, 2006; Al-Mahdawi *et al*, 2008; Kim *et al*, 2011; Kumari *et al*, 2011). The *FXN* promoter in patient-derived cells and tissues shows a less permissive configuration for transcription initiation when compared with healthy cells (Al-Mahdawi *et al*, 2008; Kumari *et al*, 2011; Chutake *et al*, 2014; Groh *et al*, 2014).

I optimised the ChIP technique (Fig. 3.2) to test for the presence of repressive epigenetic hallmarks at the *FXN-GAA-Luc* transgene. Key steps included (i) determination of appropriate crosslinking time, (ii) adjustment of sonication time to yield sheared chromatin fragments with sizes ranging from 200-800 bp, (iii) using antibodies that bind their target with high affinity, and (iv) quantification of purified DNA by real-time PCR (Carey *et al*, 2009).

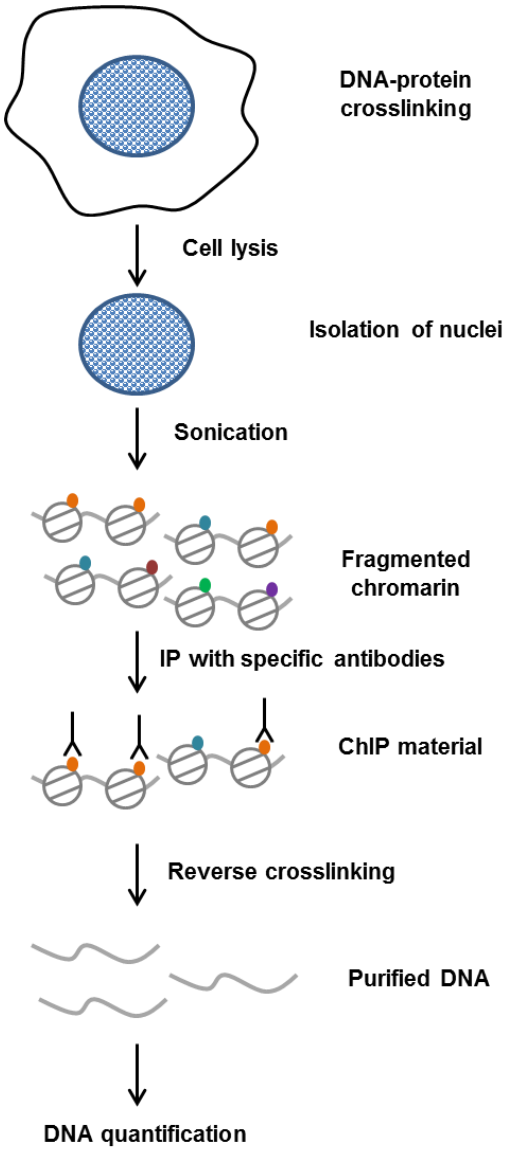

Figure 3.2 Workflow of Chromatin Immunoprecipitation (ChIP). ChIP is used to study interactions between specific proteins or modified forms of proteins and a genomic DNA region, such as the presence of histones with post-translational modifications at specific genomic locations. Formaldehyde crosslinks proteins to DNA. Nuclei are isolated and crosslinked chromatin is then sonicated to generate smaller DNA fragments. Immunoprecipitation (IP) with the desired antibodies is performed. The precipitated protein-DNA is then purified, treated with a protease, and analysed real-time PCR.

Proteins were crosslinked to chromatin to analyse histone modifications at three *FXN* regions, namely the promoter and the regions upstream and downstream of GAA repeat expansion. The best sonication conditions yielded fragments with sizes 200-800 bp and were obtained after treating $5x10^6$ *FXN-GAA-Luc* cells (10x dilution) with 1% formaldehyde for 7 minutes and performing 24 cycles of sonication (Fig. 3.3A). Immunoprecipitation of histones bound to DNA was performed using antibodies specific

for the human acetylated histones H3K9 and H4K8 and di- and trimethylated histone H3K9. Crosslinking was then reversed and DNA was quantified by real-time PCR. This analysis revealed a decrease in histone acetylation and an increase in histone methylation across the three regions in the *FXN-GAA-Luc* cell line when compared with *FXN-Luc* cells (Fig. 3.3B), as previously reported in FRDA patient brain tissue and patient-derived cell lines (Herman *et al*, 2006; Al-Mahdawi *et al*, 2008; Kim *et al*, 2011; Kumari *et al*, 2011).

I then analysed CpG methylation by bisulphite sequencing as reported in previous studies (Al-Mahdawi *et al*, 2008). Genomic DNA from *FXN-GAA-Luc* and *FXN-Luc* cells was bisulphite-treated to deaminate unmethylated cytosines in order to produce uracils in DNA. In this process, methylated cytosines are protected from the conversion to uracils. The *FXN* promoter and regions flanking the GAA repeats were amplified by PCR using bisulphite-specific primers and then sequenced to determine the locations of unmethylated cytosines and 5-methylcytosines at single-nucleotide resolution (Fig. 3.4A). The *FXN* promoter region did not present any of the 59 CpG sites methylated in either cell line. In line with these results, Greene *et al* (2007) also reported a lack of any DNA methylation in the promoter region of either healthy or FRDA patient-derived cells. However, depending on the type of samples analysed, the promoter has shown (i) incomplete methylation at specific CpG sites in the heart tissue of healthy controls, but complete methylation of those CpG sites in FRDA (Al-Mahdawi *et al*, 2008), and also (ii) no observed significant difference between healthy and FRDA clinical samples in the promoter region with CpG sites being largely unmethylated (Evans-Galea *et al*, 2012). In the region upstream of the expanded GAA repeats, I found increased DNA methylation at CpG sites 4 and 5 (Fig. 3.4B), in agreement with previously published methylation data from FRDA patients (Al-Mahdawi *et al*, 2008; Evans-Galea *et al*, 2012). Data for CpG 6 differ from the results previously found in patients (Al-Mahdawi *et al*, 2008; Greene *et al*, 2007), but are in agreement with methylation data from the brain of FRDA mice (Al-Mahdawi *et al*, 2008). In the region downstream of the GAA repeats, an increase in DNA methylation was observed at CpG sites 1, 2 and 7 (Fig. 3.4B).

Overall, since the *FXN-Luc* and *FXN-GAA-Luc* cell lines share the same integration site and only differ by the presence of an expansion of ~310 GAA repeats, these data suggest that the GAA expansion is causing the observed changes in histone acetylation/methylation and DNA methylation and inducing a closed chromatin configuration which may be linked to *FXN-GAA-Luc* repression.

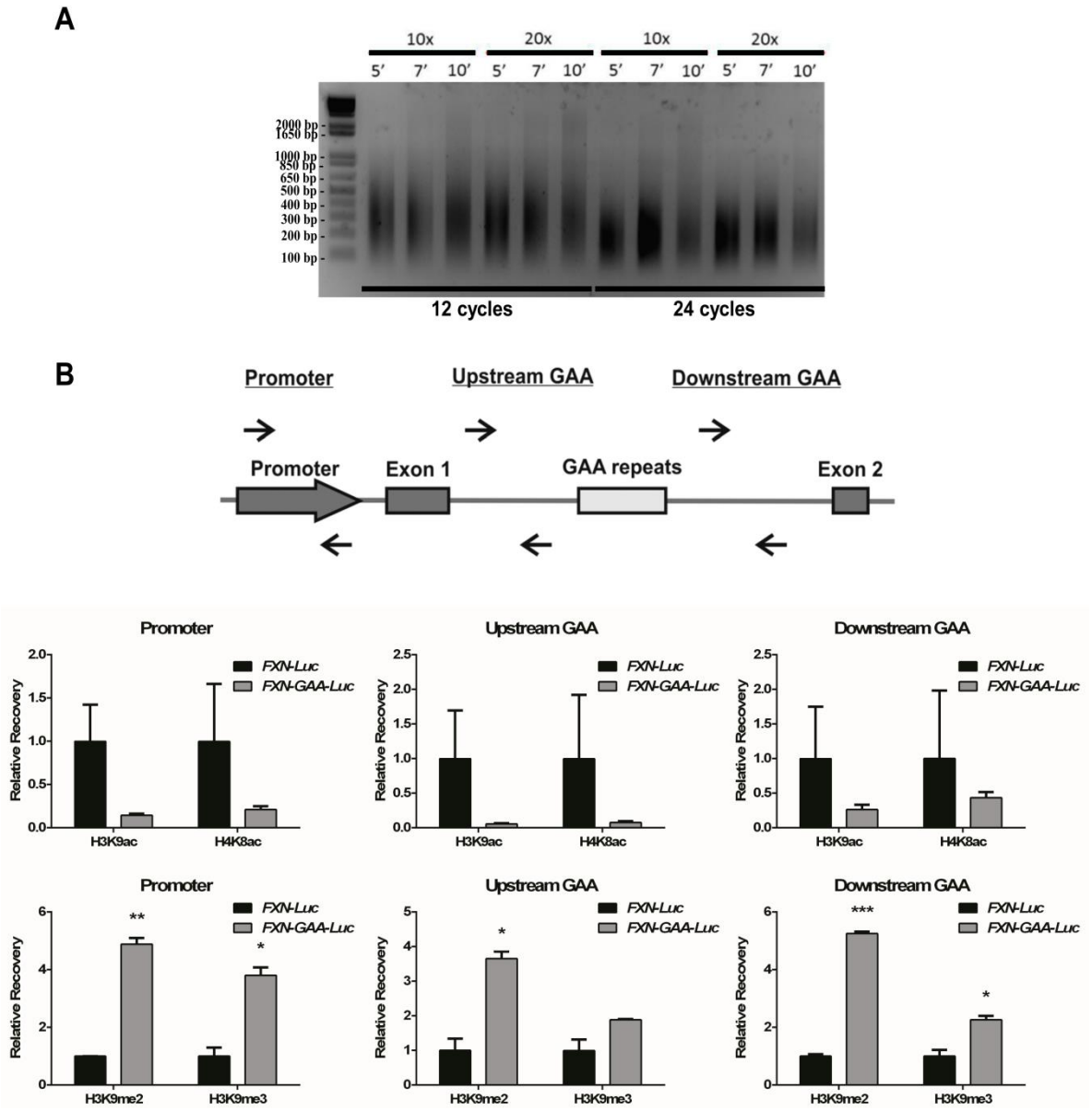

Figure 3.3 Histone modifications on the *FXN* locus. (**A**) Optimization of chromatin shearing was performed in *FXN-GAA-Luc* cells. Serial dilutions were made starting with $50x10^6$ cells and diluting 10x and 20x. Twelve and 24 cycles of sonication of 30 seconds "ON" and 30 seconds "OFF" were performed after crosslinking proteins to the DNA for 5, 7 and 10 minutes. Crosslinking was reversed overnight. (**B**) Schematic representation showing the location of primers used for DNA quantification by real-time PCR (top). ChIP was performed on *FXN-Luc* and *FXN-GAA-Luc* cells using antibodies specific for the human acetylated histones H3K9 and H4K8 (middle) and di- and trimethylated histone H3K9 (bottom). Analysis of histone modifications in the promoter and regions flanking GAA repeats showed a decrease in H3K9ac and H4K8ac levels and an increase in H3K9me2 and H3K9me3 levels in *FXN-GAA-Luc* cells. Data are mean ± SEM from two independent IPs and each IP was quantified in triplicate. **P* < 0.05, ***P* < 0.01, ****P* < 0.001, as determined by unpaired two-tailed Student's t-test.

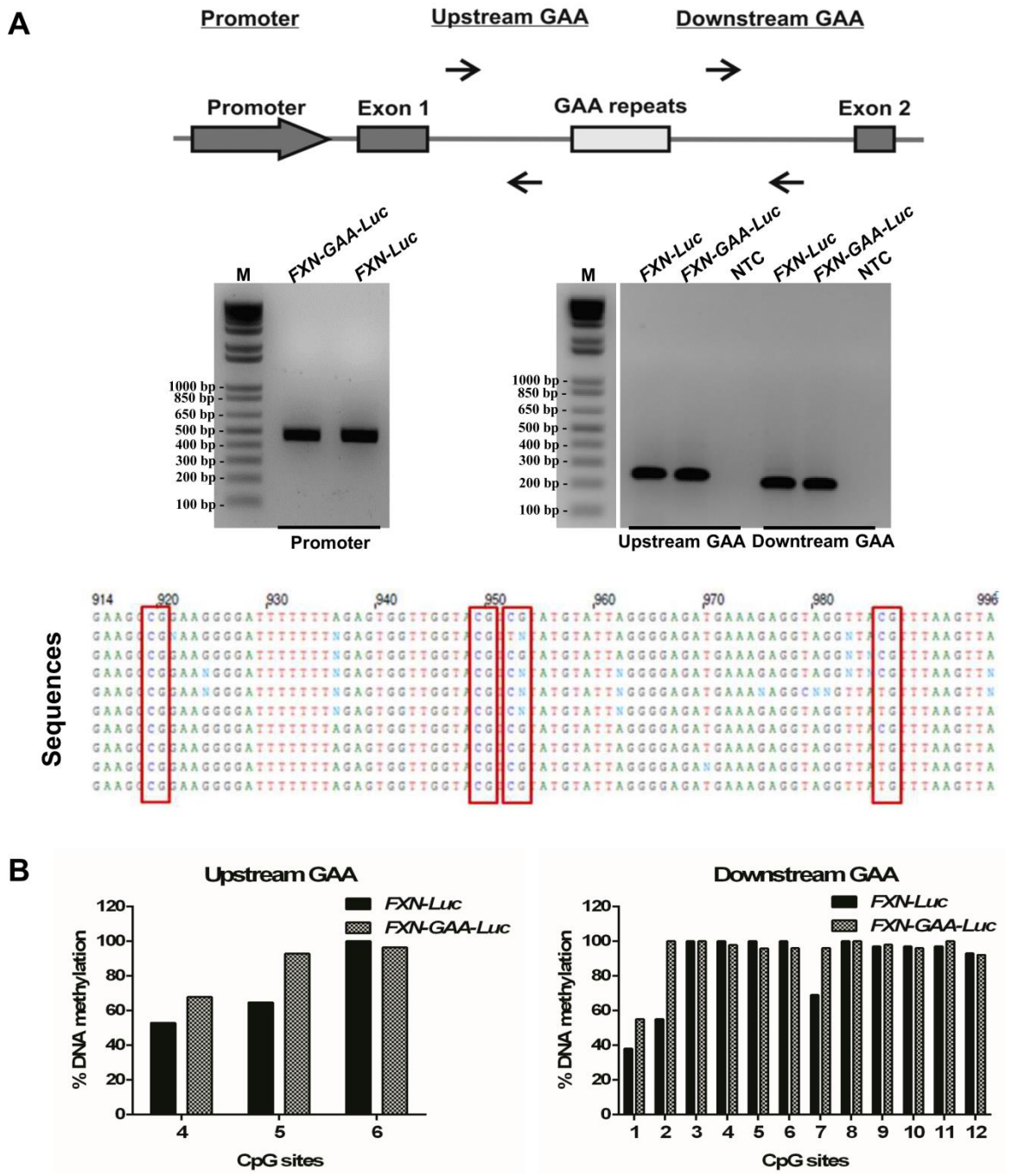

Figure 3.4 DNA methylation analysis of the *FXN* locus. (**A**) Schematic representation showing the location of the bisulphite-specific primers used in this analysis (top). Genomic DNA of the *FXN-Luc* and *FXN-GAA-Luc* cell lines was bisulphite-converted and the promoter and regions flanking the GAA repeats were amplified by nested-PCR (middle). PCR products were cloned and a total of 10 colonies were sequenced per region and for each cell line. A section of the sequencing data is shown (bottom). CpG sites are identified in red boxes. (**B**) The promoter region did not show any CpG methylation in either cell line. *FXN-GAA-Luc* cells showed an increased in DNA methylation at CpG sites 4 and 5 upstream of the GAA expansion and at CpG sites 1, 2 and 7 downstream of GAA repeats.

3.2.3 C5 restores histone acetylation levels around the GAA repeat expansion

Treatment with the HDAC inhibitors nicotinamide and 2-aminobenzamide compounds 106, 136, and 109 have been previously shown to upregulate *FXN* expression in FRDA patient cells (Xu *et al*, 2009; Chan *et al*, 2013) and mouse models (Rai *et al*, 2010; Sandi *et al*, 2011). The *FXN-Luc* and *FXN-GAA-Luc* cell lines allow the rapid quantification of *FXN* expression and are a useful tool for the identification of compounds able to increase frataxin protein levels. The GAA-expanded reporter model was used in the laboratory to screen a library of novel small molecules with potential HDAC inhibitor function and led to the identification of one molecule (C5) which upregulates *FXN* expression in FRDA patient primary cells.

To assess whether C5 acts as an HDAC inhibitor on the *FXN* gene and is able to reverse the GAA-mediated *FXN* silencing, I incubated FRDA patient-derived lymphoblastoid cells with C5 at 20 mM for 48 h and analysed histone acetylation at the *FXN* gene by ChIP (Fig. 3.5). The best sonication conditions yielded fragments with sizes 200-800 bp and were obtained after treating $2.5x10^6$ cells (20x dilution) with 1% formaldehyde for 4 minutes and performing 24 cycles of sonication (Fig. 3.5A). Analysis of H3K9 and H4K8 acetylation was performed at the *FXN* promoter and regions upstream and downstream of GAA repeats (Fig. 3.5B). Patient-derived cells showed a reduction in histone acetylation at the regions flanking the GAA expanded repeats when compared with healthy cells, in line with previous reports (Herman *et al*, 2006; Kim *et al*, 2011; Kumari *et al*, 2011), but not in the *FXN* promoter. However, reduced levels of some marks of active chromatin at the *FXN* promoter, but not others, have been reported previously (Kim *et al*, 2011; Kumari *et al*, 2011). When FRDA patient cells were incubated with C5, the FRDA histone acetylation was restored to healthy levels in the upstream and downstream regions of the GAA repeats, suggesting that C5 acts either directly or indirectly in inhibiting HDAC activity. Taken together, these data suggest C5 upregulates *FXN* expression by restoring normal histone acetylation around the GAA repeats.

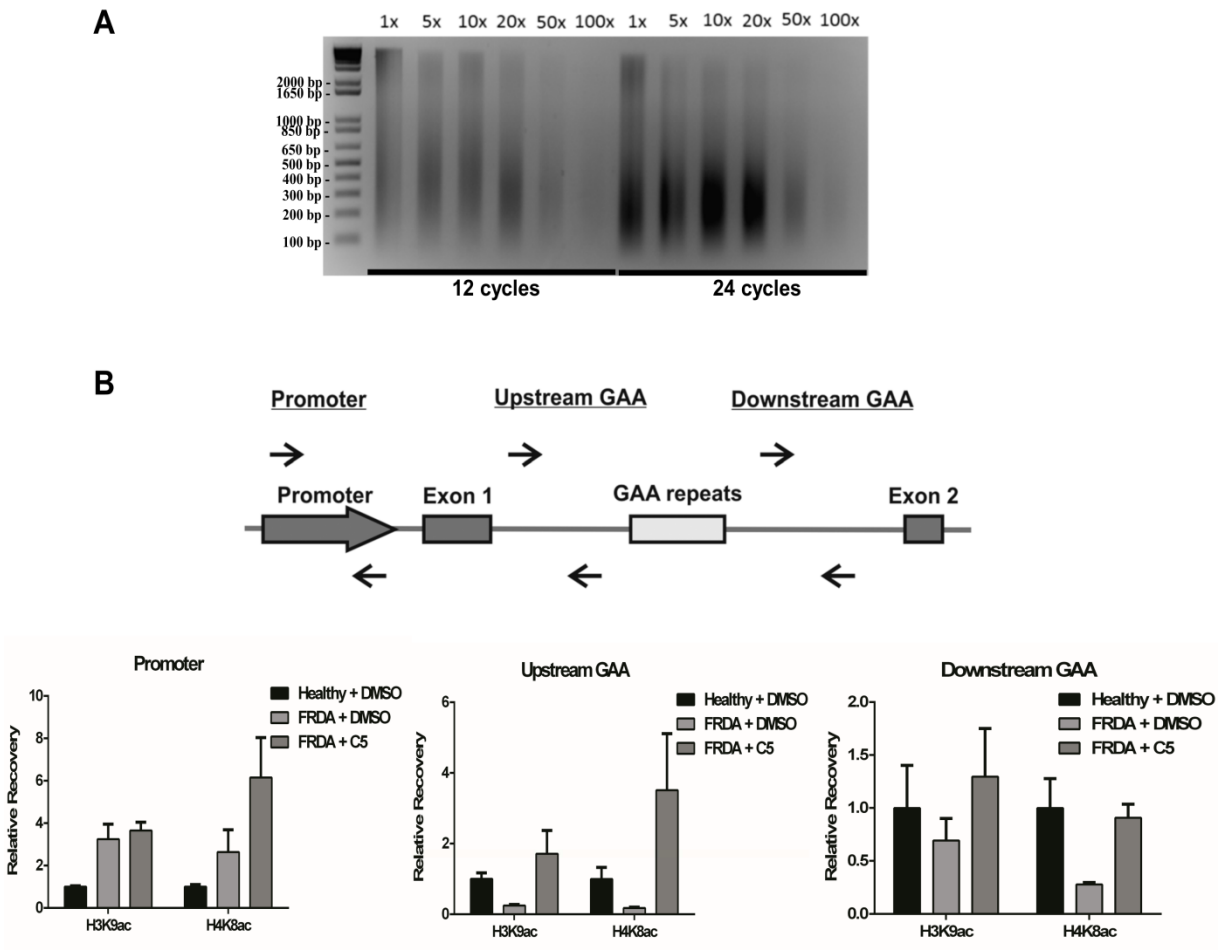

Figure 3.5 Effect of C5 on a FRDA patient-derived cell line. (**A**) Optimisation of chromatin shearing was performed using GM15850 cells. Serial dilutions were made starting with $50x10^6$ cells and diluting as follows: 1x, 5x, 10x, 20x, 50x and 100x. Twelve and 24 cycles of sonication of 30 seconds "ON" and 30 seconds "OFF" were performed for each cell line after crosslinking proteins to the DNA for 4 minutes. Crosslinking was reversed overnight. (**B**) Schematic representation showing the location of primers used for DNA quantification by real-time PCR (top). C5 increases H3K9 and H4K8 histone acetylation in FRDA lymphoblastoid cells (GM15850) to normal levels observed in healthy cells (GM15851) in the regions flanking the GAA repeats. Cells were incubated with C5 at 20 mM for 48 hours and ChIP was performed using antibodies specific for H3K9ac and H4K8ac. Data are mean ± SEM from three independent IPs and each IP was quantified in triplicate.

3.3 Discussion

Expanded GAA repeats lead to *FXN* mRNA deficiency and subsequent reduced levels of frataxin protein in FRDA. To understand the underlying pathogenesis and ultimately to design effective therapies for FRDA, it is important to generate cellular models that recapitulate the repressive hallmarks of the disease. All existing cell models present advantages and disadvantages (reviewed in Perdomini *et al*, 2013). As human primary neurons and cardiomyocytes cannot be obtained, the most relevant available cells

for studying the pathological mechanisms in FRDA are patient-derived fibroblasts and lymphocytes. These cells contain two *FXN* alleles with the entire genomic locus carrying expanded GAA repeats in their native chromatin environment. Although the mutation reduces *FXN* expression in these cells, expanded GAA repeats are relatively stable (Sharma *et al*, 2002) and the pathological biochemical phenotype is absent (Sturm *et al*, 2005). FRDA patient skin fibroblasts have been reprogrammed into iPSCs (Ku *et al*, 2010; Liu *et al*, 2011; Hick *et al*, 2013), which have been differentiated into neurons and cardiomyocytes (Liu *et al*, 2011; Hick *et al*, 2013). These iPSCs exhibit GAA repeat instability and reduced levels of *FXN* mRNA and protein, but do not show biochemical defects (Hick *et al*, 2013). Differentiated neurons and cardiomyocytes present mitochondrial dysfunction (Hick *et al*, 2013). However, differentiation protocols are not efficient and yield a mixed population of cells.

The *FXN-Luc* and *FXN-GAA-Luc* reporter cell lines are the first models that allow a direct comparison of the effect of normal or expanded GAA repeats on *FXN* expression within the genomic context of the *FXN* locus. A few reporter cell models carrying expanded GAA repeats have been described previously. However, these models only carry expanded GAA repeats within a *GFP* reporter gene lacking any *FXN* gene sequence (Soragni *et al*, 2008; Ditch *et al*, 2009) or only contain a small portion of intron 1 (Grant *et al*, 2006). The *FXN-GAA-Luc* and *FXN-Luc* cell lines carry the whole *FXN* locus with its native promoter, introns, exons and all elements required for physiological transgene expression. Furthermore, the cell model was generated by site-specific recombination to promote the integration of a single copy of pBAC-*FXN-Luc* or pBAC-*FXN-GAA-Luc* into a specific FRT (Flp recombinase target) site in chromosome 1 of HEK FRT cells, generating highly comparable cell lines only differing in the number of GAA repeats. A previous *FXN* genomic reporter system has been described previously; however, this system does not contain expanded GAA repeats (Sarsero *et al*, 2003; Li *et al*, 2013).

RNA/DNA hybrids (R-loops) have been previously shown to form on the *FXN-GAA-Luc* transgene (Groh *et al*, 2014). R-loops impair RNAPII elongation and trigger epigenetic changes, such as H3K9me2 repressive marks through recruitment of G9a methyltransferase (Groh *et al*, 2014).

Similar to other cellular models, the *FXN-GAA-Luc* reporter model exhibits limitations. Due to the difficulty in engineering large GAA repeat expansions, the *FXN-GAA-Luc* cells contains only an expansion of ~310 GAA repeats, which does not cause a severe phenotype as seen in FRDA patient cells. Nevertheless, ~310 GAA repeats cause

FRDA with a late age of onset at ~30 years old (Bidichandani *et al*, 2000). Because the endogenous *FXN* gene is present in the *FXN-GAA-Luc* cell line, the FRDA biochemical defects cannot be analysed. Moreover, this cell model does not show time-depended GAA repeat instability. However, in this regard, it would be interesting to assess the expression levels of the MMR proteins MSH2, MSH3, MSH6 or PMS2 and analyse their effect in inducing GAA expansions or contractions in these cells, like in other cell models (Ku *et al*, 2010; Du *et al*, 2012) and mouse models (Ezzatizadeh *et al*, 2012; Bourn *et al*, 2012).

The presence of ~310 GAA repeats reduces *FXN-GAA-Luc* mRNA and protein levels via a heterochromatin-mediated silencing mechanism. I show that the GAA repeat expansion decreases histone acetylation and increases histone methylation at the *FXN* promoter and regions upstream and downstream of the GAA repeats, in accordance with previous reports (Herman *et al*, 2006; Al-Mahdawi *et al*, 2008; Kim *et al*, 2011; Kumari *et al*, 2011). Furthermore, *FXN-GAA-Luc* cells show increased CpG methylation upstream and downstream of the GAA repeat expansion when compared with *FXN-Luc* cells. Upstream of the GAA repeats, I observe increased DNA methylation at CpG sites 4 and 5, in accordance with previous data obtained using post-mortem brain and heart tissues from FRDA patients (Al-Mahdawi *et al*, 2008) and peripheral blood and buccal cells from FRDA patients (Evans-Galea *et al*, 2012). The results for CpG site 6 are different from methylation data from FRDA patients (Al-Mahdawi *et al*, 2008; Greene *et al*, 2007), but are in agreement with data from brain tissue of FRDA mice (Al-Mahdawi *et al*, 2008). In the region downstream of the GAA repeats, I observed increased methylation at CpG sites 1, 2 and 7, whereas the rest of the CpG sites were completely methylated. Unlike these findings, analysis of DNA methylation in cells (Evans-Galea *et al*, 2012) and tissues (Al-Mahdawi *et al*, 2008) from FRDA patients reported a general hypomethylation in this region. A methylation pattern similar to the findings in *FXN-GAA-Luc* has been reported inYG8 mice, which do not present hypomethylation downstream of GAA repeats (Al-Mahdawi *et al*, 2008). Since the results shown here are different from data obtained from FRDA patients (Evans-Galea *et al*, 2012), who carry large GAA expansions (618.1 ± 230.7), but are similar to the findings in YG8 mice which carry shorter GAA repeats (190 + 90 GAA repeats), it is tempting to speculate that this variation in CpG methylation could be related to the size of the GAA expansion. To explain how shorter repeat expansions cause changes in DNA methylation, Al-Mahdawi and colleagues suggested that large GAA expansions increase the distance of the downstream region from a putative methylation centre at the 5′ end of the Alu sequence, located upstream of GAA repeats, thus generating

hypomethylation of the region downstream of GAA repeats when compared with healthy individuals. However, a shorter GAA expansion would decrease such distance, thereby leading to hypermethylation downstream of GAA repeats (Al-Mahdawi *et al*, 2008).

Recently, HDAC inhibitors have been tested as a potential therapy for FRDA. These compounds successfully increased *FXN* expression in FRDA lymphoblastoid cells and primary lymphocytes, KIKI mice and YG8R mice (Herman *et al*, 2006; Rai *et al*, 2008, 2010; Sandi *et al*, 2011; Chan *et al*, 2013). The *FXN-Luc* and *FXN-GAA-Luc* cell lines allow the rapid quantification of *FXN* transgene expression and are a useful tool for the identification of compounds able to increase frataxin protein levels. C5 was identified in the laboratory using this cell model. Subsequently, this compound was shown to upregulate *FXN* expression in FRDA patient primary cells by remodelling the chromatin configuration around the GAA repeats.

Data presented in this Chapter demonstrates that genomic reporter models of FRDA recapitulate the GAA repeat-mediated *FXN* repression and the epigenetic hallmarks of FRDA. Therefore, these models are an excellent tool to study the molecular mechanisms underlying FRDA as well as providing a platform for the screening of *FXN* upregulating compounds.

CHAPTER 4

Visualisation of the effect of a GAA repeat expansion on *FXN* gene nuclear localisation and expression at single-cell resolution

The results featured in this Chapter were published in the paper (Annex 2):

Silva AM, Brown JM, Buckle VJ, Wade-Martins R and Lufino MM (2015) Expanded GAA Repeats Impair *FXN* Gene Expression and Reposition the *FXN* Locus to the Nuclear Lamina in Single Cells. *Hum. Mol. Genet.* 24(12):3457-71.

Author contributions:

A.M.S, J.M.B., V.J.B., R.W.-M. and M.M.L. conceived and designed the experiments; **A.M.S** performed the experiments and analysed the data; **A.M.S** wrote the manuscript; **A.M.S**, J.M.B., V.J.B., R.W.-M. and M.M.L. revised the manuscript.

4.1 Introduction

Gene expression is dependent on the spatial and structural organisation of genes inside the nucleus and a wide range of proteins involved in transcriptional regulation. Standard methods to study gene expression – such as northern blotting, qRT-PCR and microarrays – rely on the detection of transcripts from samples containing millions of cells. These steady-state population-level measurements are useful if a considerable proportion of the cell population is actively expressing. However, if the sample is small and/or the target is present in low copies not reaching a detectable threshold, changes in expression levels can go unnoticed. Single-cell analyses in various systems have revealed a considerable cell-to-cell variation in gene expression, and this heterogeneity can be attributed to the stochastic nature of transcription (Levsky & Singer, 2003; Coulon *et al*, 2013). Given that averaged gene expression levels in cell populations do not reflect the behaviour of individual cells, methods that can detect and quantify mRNA molecules in single cells are needed. In this introduction, I will highly the experimental advantages of studying transcriptional kinetics at the single-cell level.

4.1.1 *In situ* hybridisation (ISH)

ISH commonly uses nucleic acid probes that are complementary to target RNA sequences to reveal the intracellular location of mRNAs in fixed cells (reviewed in Weil *et al*, 2010). Improvements due to technical advances typically revolve around the type of probe and labelling used to detect the signal. Initially, radiolabelled cDNA probes complementary to the appropriate target were used (Harrison *et al*, 1973). Although quantitative, detection offered low spatial resolution and the technique was laborious and time-consuming. Later, it became conventional to use probes containing modified bases that allowed the conjugation of various haptens or fluorophores. Detection could be performed indirectly through enzymes that catalyse chromogenic or fluorogenic reactions or directly by fluorescent microscopy (Raap *et al*, 1995). Indirect detection is highly sensitive due to signal amplification but the location of signal is not precisely determined. In contrast, probes labelled directly with a few fluorophores maintain spatial resolution, but sensitivity is relatively poor. High sensitivity is of paramount importance because the majority of mRNAs are often present at less than ten copies per cell (Lockhart & Winzeler, 2000).

Improvements in RNA FISH methodology were made to increase sensitivity and to allow the detection of single mRNAs (Femino, 1998). In this method, five 50-mers

oligonucleotide probes labelled with 5 fluorochromes per molecule are hybridised to each mRNA target and are easily visible as a fluorescent spot via fluorescence microscopy. The authors were able to determine the rates of transcription initiation and termination and RNA processing by targeting the probes to the *β-actin* mRNA. However, these probes are prone to hybridisation variability and self-quenching due to the presence of multiple fluorophores in the same oligonucleotide.

To overcome these issues, Raj and colleagues targeted mRNA molecules with 48 different oligonucleotides singly labelled with Alexa 594 flurophores. This method generates uniform signals that can be computationally identified to determine accurate mRNA counts, allowing simultaneous detection of single molecules of multiple mRNAs in single cells (Raj *et al*, 2008).

4.1.2 mRNA imaging in living cells

Important data regarding transcription and mRNA localisation have been harnessed from RNA FISH experiments in fixed cells. However, they show static snapshots of dynamic processes that may only be occurring in a subset of cells. Visualising specific mRNAs in living cells requires techniques to fluorescently tag mRNAs to monitor their dynamics over time.

RNA FISH can also be used in living cells, but it has its limitations (review in Sheinberger & Shav-Tal, 2013). Cells can tag short single stranded DNA oligonucleotides in the nucleus for rapid degradation, and the inability to properly wash away unbound probe leads to high background. Improvements have been made by designing degradation-resistant molecular beacons consisting of short oligonucleotide probes with a fluorophore and a quencher on opposite ends. They are delivered to cells in a quenched fluorescent state and only fluoresce when specifically hybridised to their mRNA target (Vargas *et al*, 2005). This reduces background and allows single-molecule resolution. Limitations include (i) probe delivery via microinjection, electroporation or streptolysin O permeabilisation, which can be challenging to living cells; (ii) inefficient probe hybridisation due to RNA-protein binding or to secondary RNA structures; and (iii) the need for multiple binding sites to generate a detectable signal.

A substantial advance in mRNA imaging was achieved using the MS2 system which allows the direct fluorescent tagging of mRNAs due to the specific interaction between the bacteriophage MS2 coat protein (MCP) and a unique RNA hairpin sequence (MS2-binding site (MBS)) (Bertrand *et al*, 1998). When the MBS is transcribed, it forms stem-loops that

are bound by dimers of MCP (Fig. 4.1). Inserting the MBS in tandem and co-expressing the MCP fused to a fluorescent protein (MCP-FP) enables time-lapse imaging of transcription and mRNA kinetics in living cells. This method also has its limitations: (i) the number of loops inserted into an mRNA needs to be carefully considered as to not increase the size of the RNA substantially, but still result in signal amplification; (ii) it identifies only mRNA molecules that contain the MBS originated from exogenous gene constructs; (iii) the right location for the MBS insertion is subjective; and (iv) the expression levels of the MCP-FP have to be adjusted to achieve a good signal-to-noise ratio and avoid aggregates in cells expressing the MCP-FP without the target RNA (Yunger *et al*, 2013).

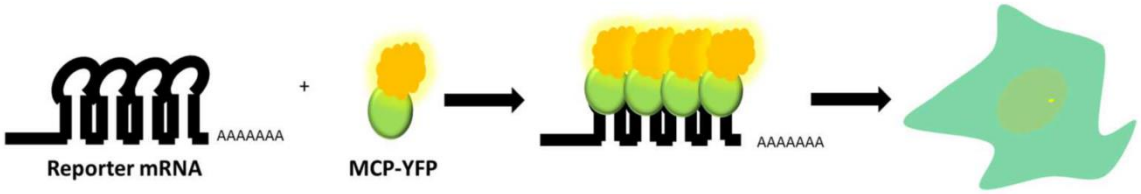

Figure 4.1 The MS2 system allows real-time imaging of RNA synthesis at a single-cell level. In this system, when a gene carrying the DNA sequence derived from the bacteriophage MS2 (MBS) is transcribed, the MBS folds into stem-loops that are bound with high affinity by the MCP. Co-expression of the MBS-gene construct with the MCP fused to a fluorescent protein (MCP-YFP) generates a nascent RNA reporter system. Sharp fluorescent spots inside the nucleus indicate active transcription sites above the unbound MCP-YFP background.

Similar systems exist that use the same principle of coat proteins binding hairpin sequences. The U1A mRNA system is limited to non-mammalian cells. It uses the human U1A protein, a component of the spliceosomal U1 small nuclear ribonucleoprotein, and a specific RNA hairpin (Brodsky & Silver, 2002). Simultaneous imaging of two mRNA species is possible by using the binding of the phage PP7 coat protein (PCP) to its RNA hairpin (Hocine *et al*, 2013), or the λ-phage N-protein–boxB system (Lange *et al*, 2008), in parallel with the MS2 system.

4.1.2.1 The MS2 system: improvements and applications

Several studies have used the MS2 system to study different aspects of mRNA dynamics in several models, for example transcription kinetics were examined in mammalian cells (Janicki *et al*, 2004; Darzacq *et al*, 2007; Yunger *et al*, 2010; Ben-Ari *et*

al, 2010), *E. coli* (Golding *et al*, 2005), *Dictyostelium* (Chubb *et al*, 2006; Muramoto *et al*, 2012) and yeasts (Zenklusen *et al*, 2008); the dynamic movement of mRNA-protein complexes was monitor in either the nucleoplasm (Shav-Tal *et al*, 2004) or the cytoplasm (Fusco *et al*, 2003) of mammalian cells; mRNA localisation was analysed in yeasts (Bertrand *et al*, 1998), mice (Lionnet *et al*, 2011) and *Drosophila melanogaster* (Zimyanin *et al*, 2008).

The first studies of transcription in live mammalian cells combined the *lacO-lacI* and the MS2 systems together with fluorescent labelling of the protein product (Janicki *et al*, 2004; Darzacq *et al*, 2007; Ben-Ari *et al*, 2010). Using this highly sensitive system composed of a 200 copy transgene array, the authors were able to correlate changes in chromatin structure with the progression of transcriptional activation and protein production in single living cells. Photobleaching and photoactivation of MCP-FP bound to nascent mRNAs provided sensitive RNAPII elongation measurements. Polymerases elongated at 4.3 kb/min and entered a paused state for unexpectedly long times.

Using models with transgenes in tandem amplifies the mRNA fluorescent signal to highly detectable levels. Given the variability in gene expression due to the probability of a gene being switched on and off, these models provide only an averaged transcriptional output from many copies of the same gene and cannot yield a true picture of the endogenous state of single-copy genes. A major concern when imaging single genes that are expressing is whether the signal will be detectable in the nucleus of mammalian cells, given that the signal-to-noise ratio is much lower compared with using multiple genes.

Integration of the MBS into specific single-copy genes of *Dictyostelium* revealed the pulsatile nature of transcription, where the frequency, intensity and duration of pulses were measured (Chubb *et al*, 2006). Each gene displayed its own transcriptional signature. While developmental genes showed a binary pulsing behaviour controlled at the level of transcription initiation, housekeeping genes strongly modulate pulsing strength during development allowing integration of global physiological cues reaching the cell (Muramoto *et al*, 2012).

A cell and a mouse model have been generated to follow and quantify expression of single genes in mammalian cells. Yunger and colleagues generated comparable stable clonal cell lines containing single copies of either the human cyclin D1 gene driven by its promoter or a CMV promoter using HEK-293 cells with a single FRT genomic locus for Flp-In homologous recombination (Yunger *et al*, 2010). The endogenous promoter had alternating periods of gene activity as seen in transcriptional bursting, and RNA FISH data

confirmed cell-to-cell variability in mammalian cells (Raj *et al*, 2006). In contrast, the viral promoter had higher and constant expression.

The first mammalian transgenic animal for imaging mRNAs in living tissue was a knock-in mouse containing the MBS inserted into the 3′ UTR of the *β-actin* mRNA, allowing its imaging in every cell of the animal (Lionnet *et al*, 2011). The ability to detect an active *β-actin* allele was confirmed by RNA FISH and by MCP-FP labelling. Subsequently, a transgenic mouse expressing MCP-GFP was crossed with the *β-actin–MBS* mouse, generating mice with an MS2 tag on both *β-actin* alleles and expressing the MCP. This mouse model allowed for the first time the global fluorescent labelling of an endogenous mRNA in all cells and tissues of an organism (Park *et al*, 2014).

Besides the traditional labelling of single mRNA molecules using the MS2 system, improvements and other applications can be performed (Fig. 4.2). By using two systems in parallel, for instance MS2 and PP7, two different mRNA species or two regions of the same mRNA can be detected (Hocine *et al*, 2013). To improve signal-to-noise ratio, the bimolecular fluorescence complementation system can be used. This system uses coat proteins fused to split fluorescent protein that only become fluorescent when bound to the mRNA stem-loops (Wu *et al*, 2014). In addition, the interactions of coat proteins with mRNAs can be used to alter their intracellular localisation (Katz *et al*, 2012), perform affinity purification, tethering proteins to the mRNA (Long *et al*, 2000), and simultaneously detect mRNAs and their protein products (Haim-Vilmovsky *et al*, 2011).

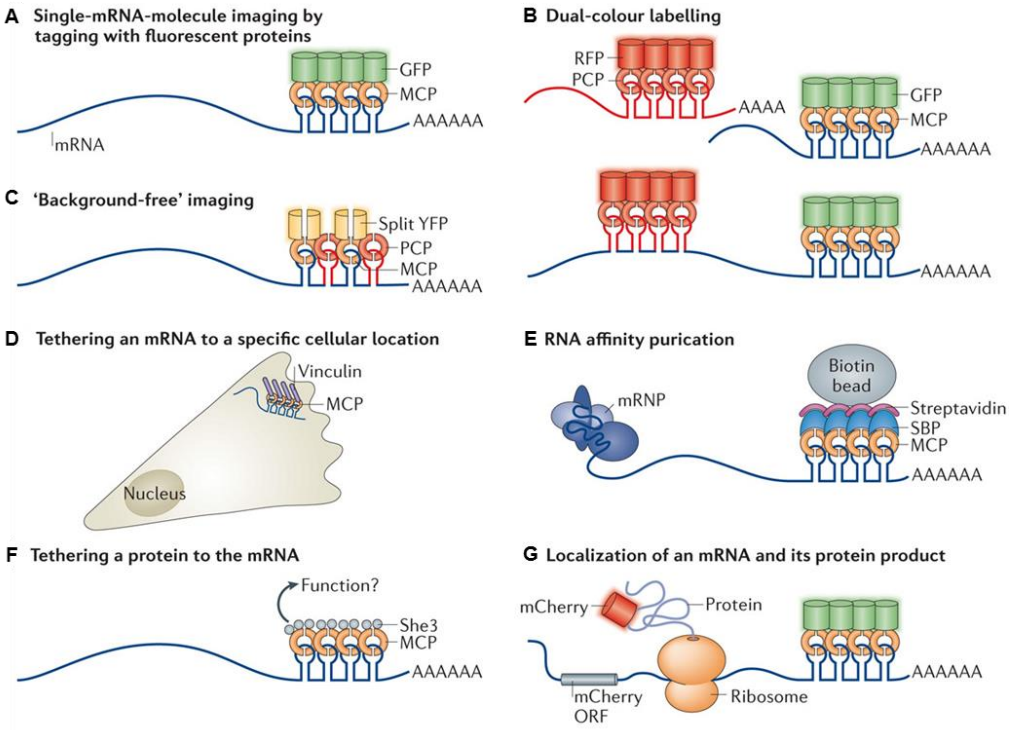

Figure 4.2 Traditional and novel uses of MS2-like systems to investigate mRNA biology. (**A**) Binding of MCP-GFP to the MBS allows single-molecule mRNA imaging. (**B**) Using two different labelling systems, two mRNAs (top) or two different parts of the same mRNA are tagged, thereby allowing the imaging of two different mRNAs in the same cell or the analysis of RNA dynamics, such as transcription, nuclear export and degradation. (**C**) To reduce background fluorescence, two different labelling systems (MS2 and PP7) can be used. Each RNA-binding protein (MCP and PCP) is fused to one half of a split fluorescent protein (YFP). Signal is obtained when both MCP and PCP are bound to the mRNA and the YFP becomes functional. (**D**) The tethering of an mRNA to a specific cellular location or structure is carried out by fusing MCP to a protein (in this case, vinculin) with specific subcellular localisation. (**E**) In RNA affinity purification, the MCP is fused to a unique epitope — for example, streptavidin-binding protein (SBP) — which mediates the affinity purification of the RNA together with protein complexes that might bind to it by using streptavidin and biotin beads. (**F**) A specific protein (for example, the transporter She3) can be tethered to an mRNA through the MCP, allowing the analysis of the RNA-protein interaction. (**G**) Simultaneous localisation of the mRNA and its protein product can be achieved by fusing the gene to an mCherry gene and cloning the MBS into its 3′ UTR (adapted from Buxbaum *et al*, 2015).

4.1.3 Aims of this Chapter

Recent studies show that expanded GAA repeats trigger R-loop formation (Groh *et al*, 2014) and locally recruit HP1 (De Biase *et al*, 2009), G9a methyltransferase (Groh *et al*, 2014) and HDACs (Herman *et al*, 2006; Chan *et al*, 2013), causing epigenetic changes that contribute to *FXN* transcriptional dysregulation. However, these studies only provide the probable state of the *FXN* gene as these observations come from experiments in which

the outputs of bulk cell cultures are averaged. A dissection of the silencing mechanism in FRDA *in situ*, in which *FXN* localisation and expression are quantified at single-cell level, is crucial to improve our understanding of the underlying pathogenesis and ultimately to design effective therapies for FRDA.

Here, I developed a novel human cell model to visualise and analyse *FXN* repression in both fixed and living single cells using the MS2 system. I show that GAA-expanded repeats decrease the number of *FXN* mRNA molecules, slow transcription, and increase *FXN* localisation at the NL.

4.2 Results

4.2.1 Construction of pBAC-*FXN-MS2-Luc* and pBAC-*FXN-GAA-MS2-Luc* vectors by homologous recombination

The pBAC-*FXN-Luc* and pBAC-*FXN-GAA-Luc* genomic DNA reporter vectors – previously used in the generation of the *FXN-Luc* and *FXN-GAA-Luc* cell lines described in Chapter 3 – were modified to develop a cell model that enables real-time visualisation of *FXN* mRNA to investigate the link between *FXN* localisation and expression at the single-cell level. These BAC vectors carry the whole 80 kb *FXN* locus with its native promoter, including exons 1 to 5b and all regulatory elements necessary to achieve physiologically-relevant *FXN* expression, a luciferase reporter gene in exon 5a and the presence of either 6 GAA repeats (pBAC-*FXN-Luc*) or a ~310 GAA repeat expansion (pBAC-*FXN-GAA-Luc*) in intron 1. A two-step selection/counter-selection homologous recombination approach in *E. coli* (Zhang *et al*, 2000) was used to insert 24 repeats of the MBS into exon 2 in a precise and specific manner (Fig. 4.3). This allows fluorescent labelling of the transgenic *FXN* mRNA and quantification of the effect of the expansion on *FXN* transcription when co-expressed with the MCP-YFP fusion protein. Given that the mature human frataxin protein is encoded by amino acids 81-210 (Schmucker *et al*, 2008), the MBS was inserted upstream of the cleavage site, allowing the production of a mature frataxin-luciferase fusion protein without the MBS.

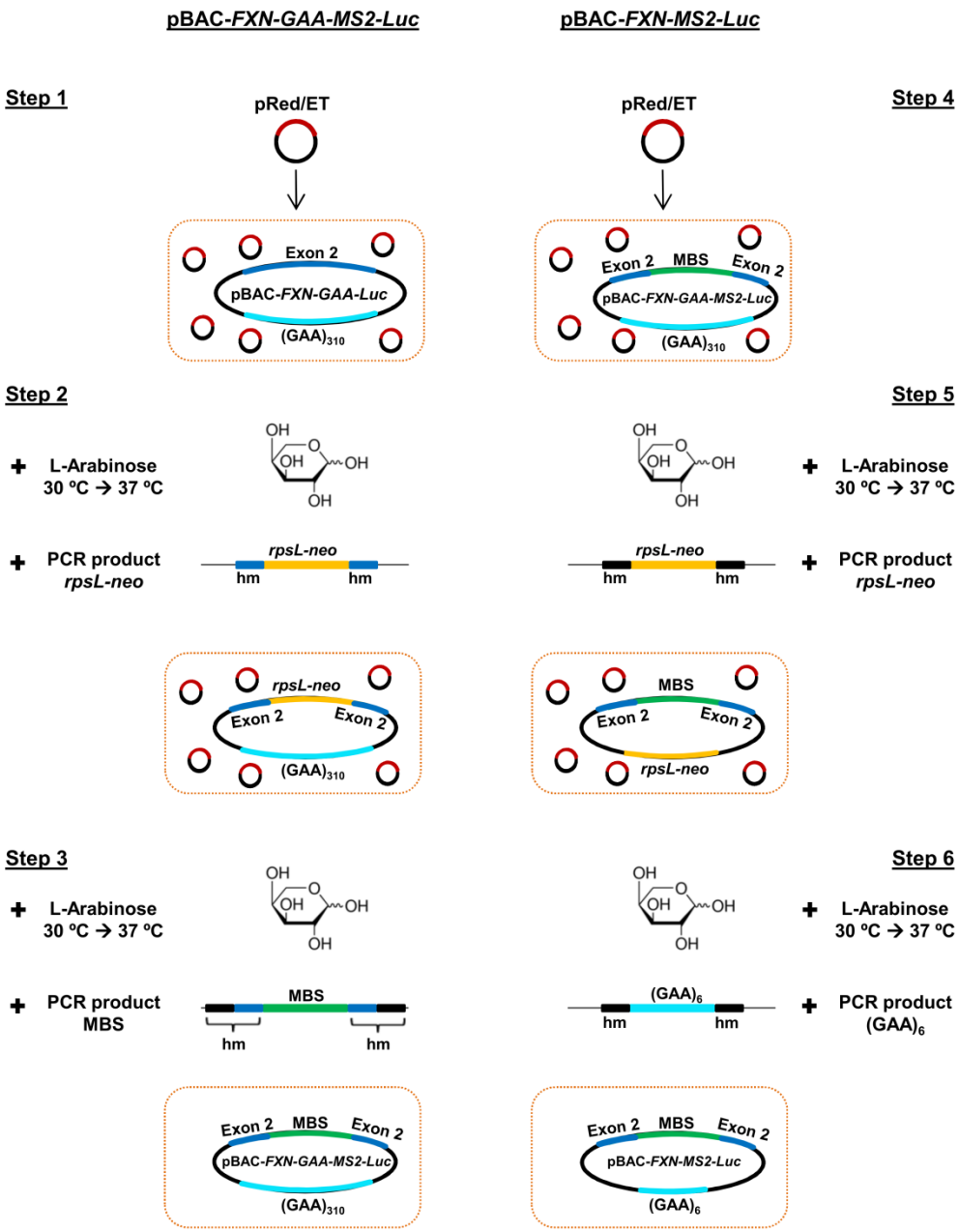

Figure 4.3 Experimental outline for vector construction of pBAC-*FXN-GAA-MS2-Luc* and pBAC-*FXN-MS2-Luc* using a selection/counter-selection homologous recombination approach in *E.coli*.
(Step 1 and 4) The *E. coli* strain, carrying the BAC to be modified, is transformed with pRedET, which mediates recombination. pRedET expresses the λ- phage *redγβα* operon under the control of the arabinose-inducible pBAD promoter. **(Step 2 and 5)** pRedET is induced by addition of L-arabinose and a temperature shift from 30°C to 37°C. After induction, cells are electroporated with the linear *rpsL-neo* counter-selection/selection cassette (PCR product) flanked by homology arms (hm). Red/ET recombination inserts *rpsL-neo* into the target locus. Only colonies carrying the modified BAC are kanamycin resistant and streptomycin sensitive. pRedET persists in cells grown at 30°C. **(Step 3 and 6)** pRedET is induced by L-arabinose and a temperature shift from 30 °C to 37 °C. After induction, cells are electroporated with the PCR product containing the MBS (step 3) or $(GAA)_6$ (Step 6), flanked by homology arms. Red/ET recombination replaces the *rpsL-neo* cassette by the non-selectable DNA. Only colonies which lost the selection/counter-selection cassette are streptomycin resistant. pBAC-*FXN-GAA-MS2-Luc* was generated from pBAC-*FXN-GAA-Luc* (steps 1-3). pBAC-*FXN-GAA-MS2-Luc* was used to generated pBAC-*FXN-MS2-Luc* (steps 4-6).

The pBAC-*FXN-GAA-MS2-Luc* was engineered from the pBAC-*FXN-GAA-Luc* vector. In the first step, a PCR product containing the *rpsL-neo* cassette flanked by 55 bp homology arms to either side of exon 2 was used to insert the cassette (Fig. 4.4A). After purification, the PCR product (Fig. 4.4B) was transformed into bacteria containing the pBAC-*FXN-GAA-Luc* construct. Successful recombinants were resistant to chloramphenicol (from the BAC vector) and kanamycin (from *rpsL-neo*), but sensitive to streptomycin (from *rpsL-neo*) (Fig. 4.4.C). Correct insertion of the *rpsL-neo* cassette was confirmed by colony-PCR (Fig. 4.4D), PCR (Fig. 4.4E) and restriction enzyme digestion followed by pulsed-field gel electrophoresis (Fig. 4.4G). Preservation of the GAA repeat expansion was assessed by colony-PCR and PCR (Fig. 4.4D and F). Before the beginning of the second step of homologous recombination, persistence of pRed/ET in successful recombinants was analysed by restriction enzyme digestion (Fig. 4.4H).

The replacement of the *rpsL-neo* cassette in the second step of recombination by 24 MBS in exon 2 was performed as follows. A plasmid containing exon 2 with an engineered *BamH*I restriction site and long flanking homology arms (147 bp upstream and 172 bp downstream of exon 2) was used to increase the efficiency of recombination. The 24 MBS were excised by *BamH*I/*Bgl*II digestion from pCR4-24XMS2SL-stable (Bertrand *et al*, 1998) and inserted in exon 2 at the *BamH*I site (Exon2-MBS) (Fig. 4.5). The *rpsL-neo* cassette was replaced with a PCR product containing the Exon2-MBS sequence flanked by homology arms, generating pBAC-*FXN-GAA-MS2-Luc* (Fig. 4.6). Successful recombinants were resistant to chloramphenicol (from the BAC vector) and streptomycin (from replacing *rpsL-neo*). Correct insertion of the 24 MBS cassette was confirmed by colony-PCR, PCR, restriction enzyme digestion followed by pulsed-field gel electrophoresis, and sequencing. Preservation of the GAA repeat expansion was assessed by colony-PCR and PCR (Fig 4.6D-F).

To generate pBAC-*FXN-MS2-Luc*, the ~310 GAA repeats in intron 1 of pBAC-*FXN-GAA-MS2-Luc* were replaced by the *rpsL-neo* cassette amplified with primers carrying 58 bp homology arms to sequences immediately upstream and downstream of GAA repeats (Fig. 4.7). The *rpsL-neo* was then replaced with a PCR product containing 6 GAA repeats amplified from the genomic DNA of a healthy-derived lymphoblastoid cell line (GM15851) (Fig. 4.8). Successful construction was confirmed by colony-PCR, restriction enzyme digestion followed by pulsed-field gel electrophoresis, and sequencing.

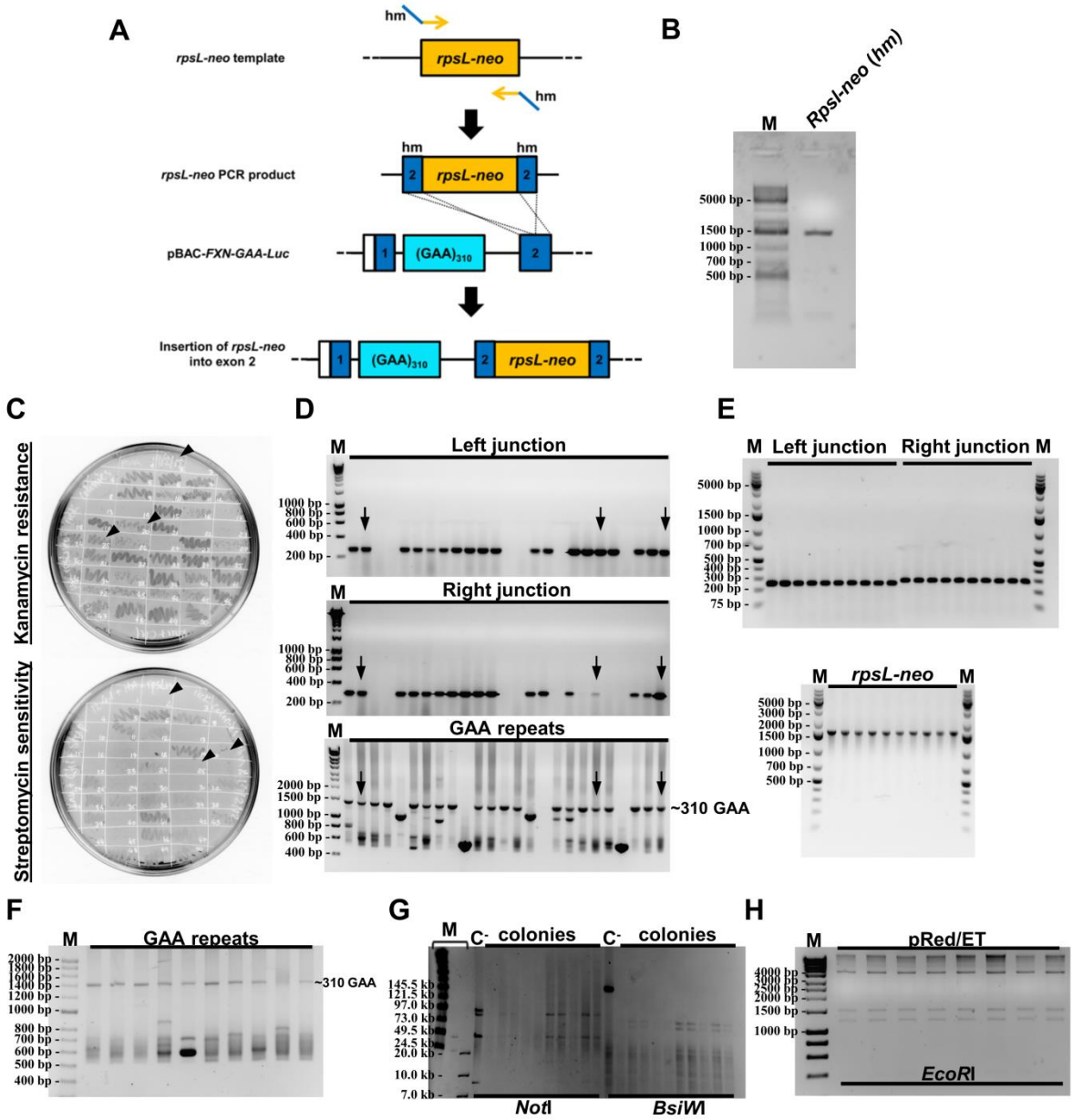

Figure 4.4 Insertion of *rpsL-neo* into exon 2 of the *FXN-GAA-Luc* gene by homologous recombination. **(A**) Schematic representation of insertion of the *rpsL-neo* cassette into exon 2 of the *FXN-GAA-Luc*. The *rpsL-neo* cassette was amplified by PCR using long primers containing homology sequences (hm) to exon 2 in pBAC-*FXN-GAA-Luc*. **(B)** The PCR product containing the *rpsL-neo* cassette flanked by 55 bp homology arms to either side of exon 2 was purified and transformed into bacteria containing the pBAC-*FXN-GAA-Luc* construct. **(C-D)** Fifty colonies were tested for: resistance to chloramphenicol and kanamycin (C – top) and sensitivity to streptomycin (C –bottom); presence of left (D – top) and right *rpsL-neo* junctions (D – middle) and full size of the GAA repeat expansion (D – bottom) by colony-PCR (only results from the first 25 colonies screened are shown). Arrows indicate successful recombinants with all the correct elements. **(E-G)** Ten successful recombinants were further tested by PCR of the left, right junctions and whole *rpsL-neo* (E) and presence of ~310 GAA repeats (F), BAC integrity was analysed by *Not*I and *BsiW*I digestion (G) C^- is pBAC-*FXN-GAA-Luc*. **(H)** Persistence of pRed/ET was analysed by *EcoR*I digestion.

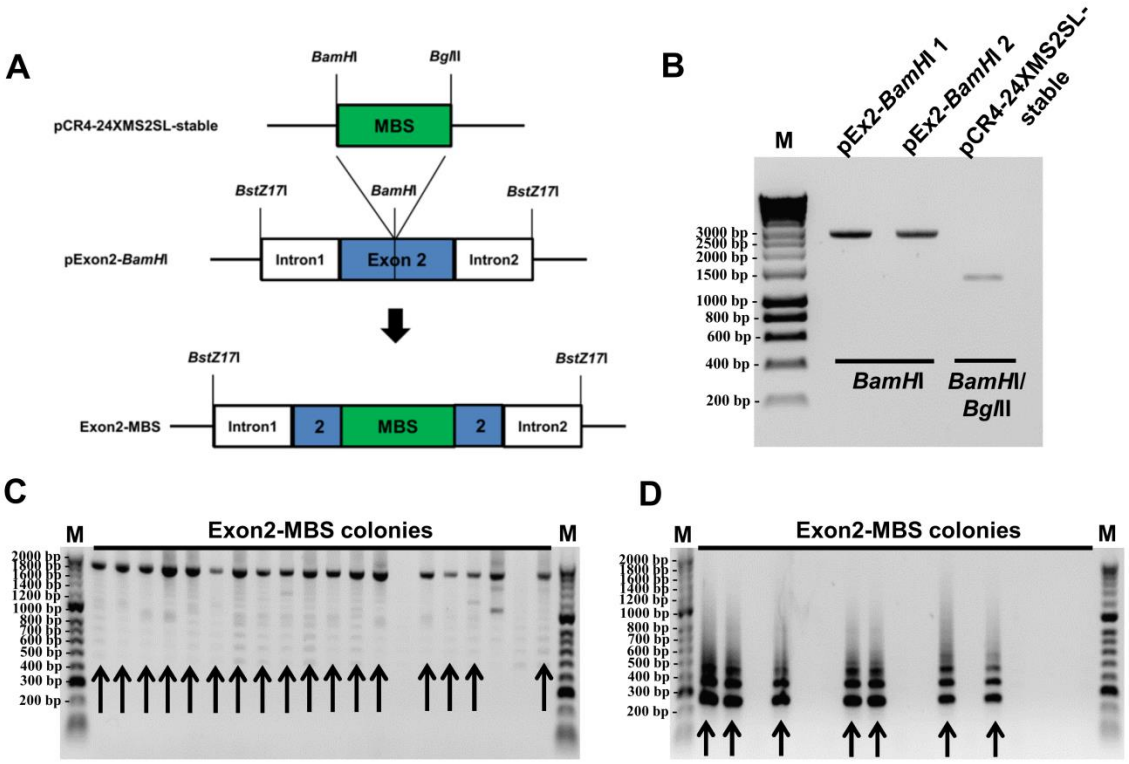

Figure 4.5 Construction of Exon2-MBS for homologous recombination. (**A**) Schematic representation of Exon2-MBS construction. pExon2-BamHI carries *FXN* exon 2 with an engineered *BamH*I restriction site and flanking homology arms (147 bp upstream and 172 bp downstream of exon 2) containing 3' *FXN* intron 1 and 5' intron 2, respectively, for subsequent homologous recombination. The 24 MBS were excised by *BamH*I/*Bgl*II digestion from pCR4-24XMS2SL-stable and inserted in exon 2 at the *BamH*I site (Exon2-MBS). (**B**) To insert the MBS cassete into the *BamHI* site of pExon2-*BamH*I, pExon2-*BamH*I was linearised by *BamH*I digestion and the MBS was excised from pCR4-24XMS2SL-stable by B*amH*I/*Bgl*II digestion. (**C**) Successful insertion of the MBS cassette was confirmed by colony-PCR. (**D**) Correct orientation of the MBS cassette was confirmed by colony-PCR. Several bands appear as the reverse primer binds to the repeated sequence of the MBS only when the insert is in the correct orientation. Arrows are positive colonies.

Cre-loxP mediated retrofitting of pBAC-*FXN-MS2-Luc* and pBAC-*FXN-GAA-MS2-Luc* vectors to pH-FRT-Hy was performed as previously described (Lufino *et al*, 2007). The pH-FRT-Hy plasmid contains the Flp-In promoter-less hygromycin cassette necessary to generate stable human clonal cell lines by site-specific integration of BAC vectors using the Flp-In system and a *lacZ* gene (Fig 4.9A). Successful recombinants were confirmed by restriction enzyme digestion followed by pulsed-field gel electrophoresis (Fig. 4.9B).

Finally, the integrity of all BAC constructs was confirmed by PCR and restriction enzyme digestion followed by pulsed-field gel electrophoresis, specifically the presence of all exons and full length GAA repeats and MBS sequences (Fig. 4.10).

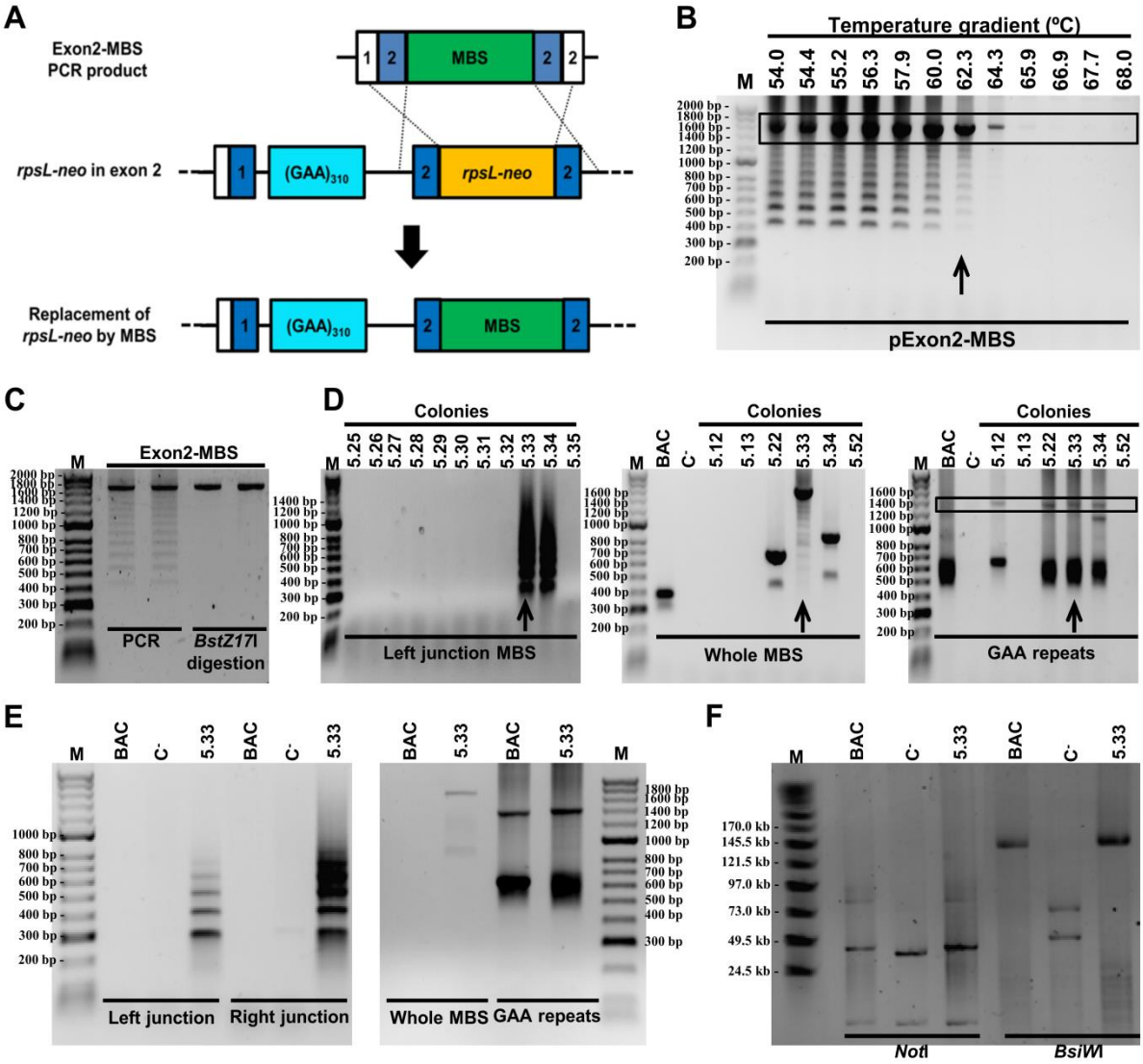

Figure 4.6 Replacement of the *rpsL-neo* cassette by 24 MBS in the second round of homologous recombination. (**A**) Schematic representation of insertion of 24 MBS into exon 2 of the pBAC-*FXN-GAA-Luc.* The *rpsL-neo* cassette was replaced by a PCR product with 24 MBS flanked by long homology arms containing exon 2, 3' *FXN* intron 1 and 5' intron 2. (**B**) Amplification of Exon2-MBS by PCR was optimised to obtain the full PCR product for homologous recombination. Arrow indicates the optimal annealing temperature. (**C**) Two strategies for homologous recombination were followed: Exon2-MBS fragment was obtained by PCR following optimised conditions as in (A) and *BstZ17*I digestion. Fragments were purified and transformed into bacteria containing the pBAC-*FXN-GAA-rspLneo-Luc* construct obtained in Figure 4.4. (**D**) One thousand and five hundred colonies resistant to chloramphenicol and streptomycin were tested for the presence of the left MBS junction by colony-PCR (left). Positive colonies were then tested for the full length MBS (middle) and GAA repeats (right) by colony-PCR. Arrow indicates the only successful recombinant (colony 5.33) with all the correct elements. BAC, pBAC-*FXN-GAA-Luc*; C^-, pBAC-*FXN-GAA-rspLneo-Luc*. (**E**) Colony 5.33 was further tested by PCR of the left, right junctions and whole MBS, and presence of ~310 GAA repeats. (**F**) BAC integrity was analysed by *Not*I and *BsiW*I digestion.

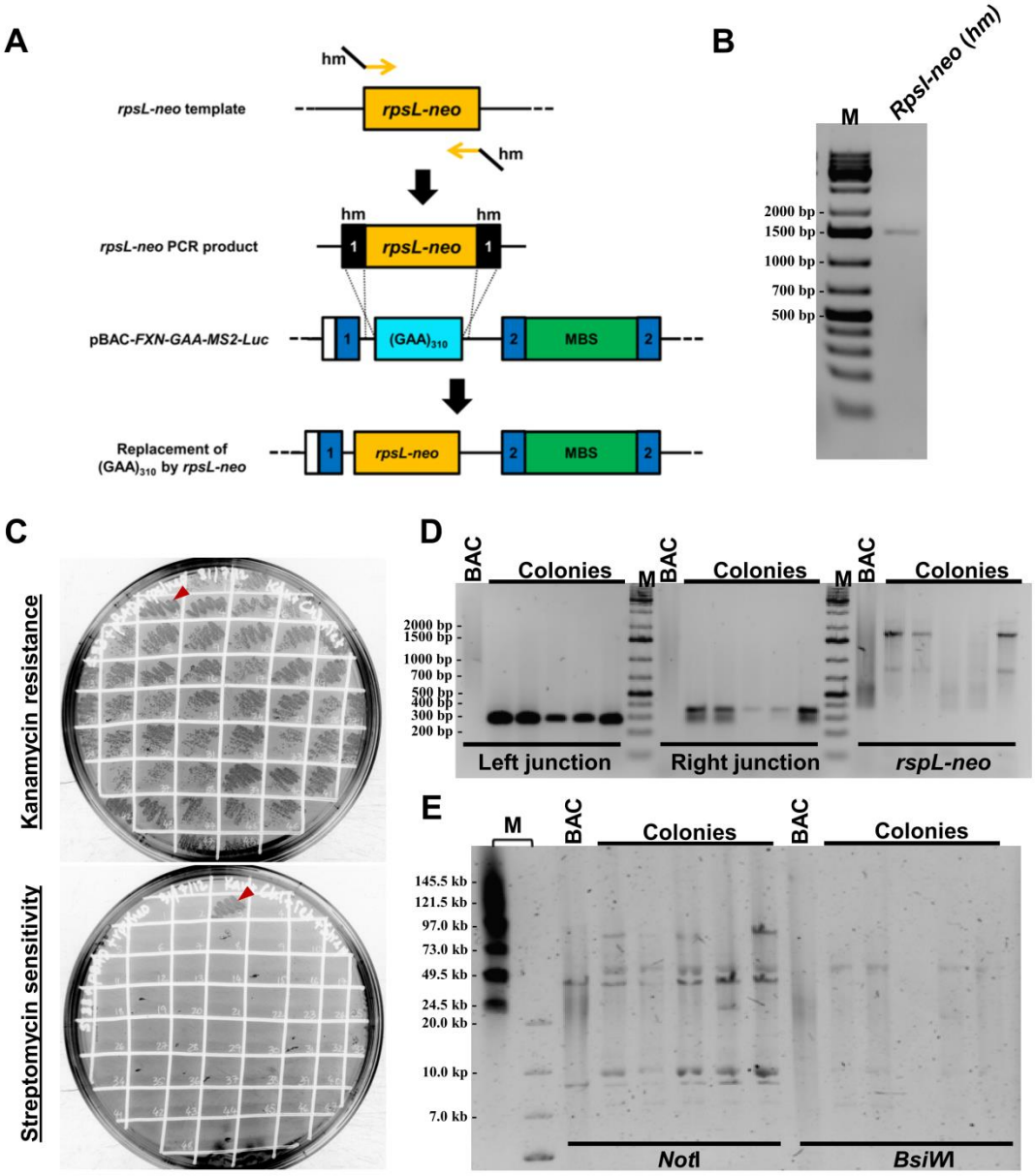

Figure 4.7 Insertion of *rpsL-neo* into intron 1 of the *FXN-GAA-MS2-Luc* gene to generate *FXN-MS2-Luc* by homologous recombination. (**A**) Schematic representation of insertion of the *rpsL-neo* cassette into intron 1 of the *FXN-GAA-Luc*. The *rpsL-neo* cassette was amplified by PCR using long primers containing homology sequences (hm) to the regions flanking the GAA repeat sequence in pBAC-*FXN-GAA-Luc*. (**B**) The PCR product containing the *rpsL-neo* cassette flanked by 58 bp homology sequences to the regions immediately upstream and downstream of GAA repeats was purified and transformed into bacteria containing the pBAC-*FXN-GAA-MS2-Luc* construct. (**C**) Forty-eight colonies were tested for resistance to chloramphenicol and kanamycin (top) and sensitivity to streptomycin (bottom). Red arrows indicate a colony resistant to streptomycin which probably carries a mutated *rpsL-neo* cassette. The remaining colonies carry a functional *rpsL-neo* cassette. (**D**) Presence of left (left), right *rpsL-neo* junctions (middle) and full size *rpsL-neo* cassette. (right) was analysed by colony-PCR. (**E**) BAC integrity was analysed by *Not*I and *BsiW*I digestion. BAC is pBAC-*FXN-GAA-MS2-Luc*.

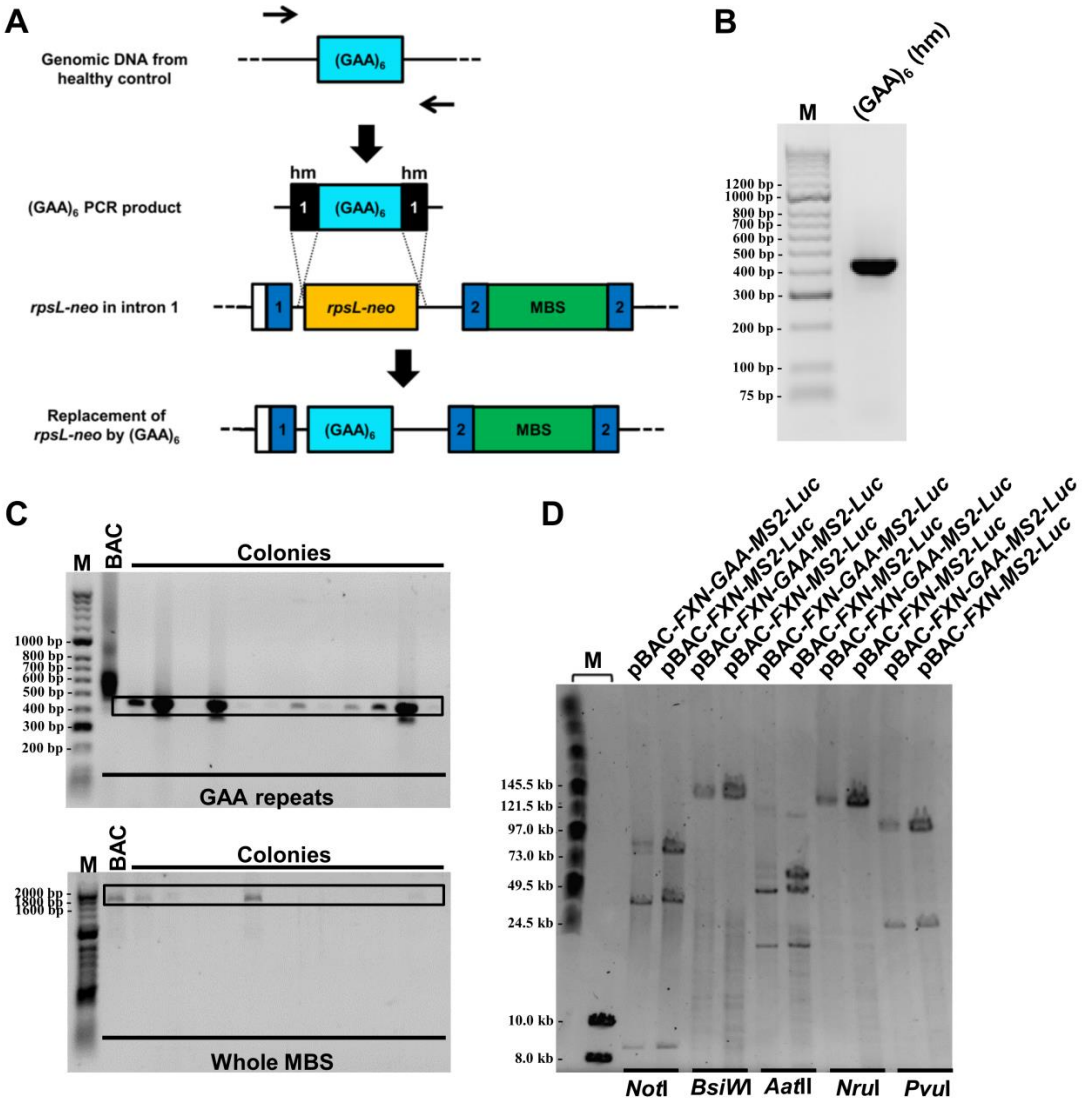

Figure 4.8 Replacement of the *rpsL-neo* cassette by 6 GAA repeats in the second round of homologous recombination. (**A**) Schematic representation of insertion of 6 GAA repeats into intron 1 of the *FXN-GAA-MS2-Luc* to generate *FXN-MS2-Luc*. The *rpsL-neo* cassette was replaced by a PCR product with 6 GAA repeats flanked by long homology arms. (**B**) The PCR product containing 6 GAA repeats flanked by homology sequences to the regions immediately upstream (195 bp) and downstream (225 bp) of GAA repeats. The fragments were purified and transformed into bacteria containing the pBAC-*FXN-rspLneo-MS2-Luc* construct obtained in Figure 4.7. (**C**) Colonies resistant to chloramphenicol and streptomycin were tested for the presence of GAA repeats (top) and full length MBS (bottom) by colony-PCR. Positive colonies were then tested for the full length MBS (middle) and GAA repeats (right) by colony-PCR. BAC, pBAC-*FXN-GAA-MS2-Luc*. (**D**) BAC integrity was analysed by *Not*I, *BsiW*I, *Aat*II, *Nru*I and *Pvu*I digestion.

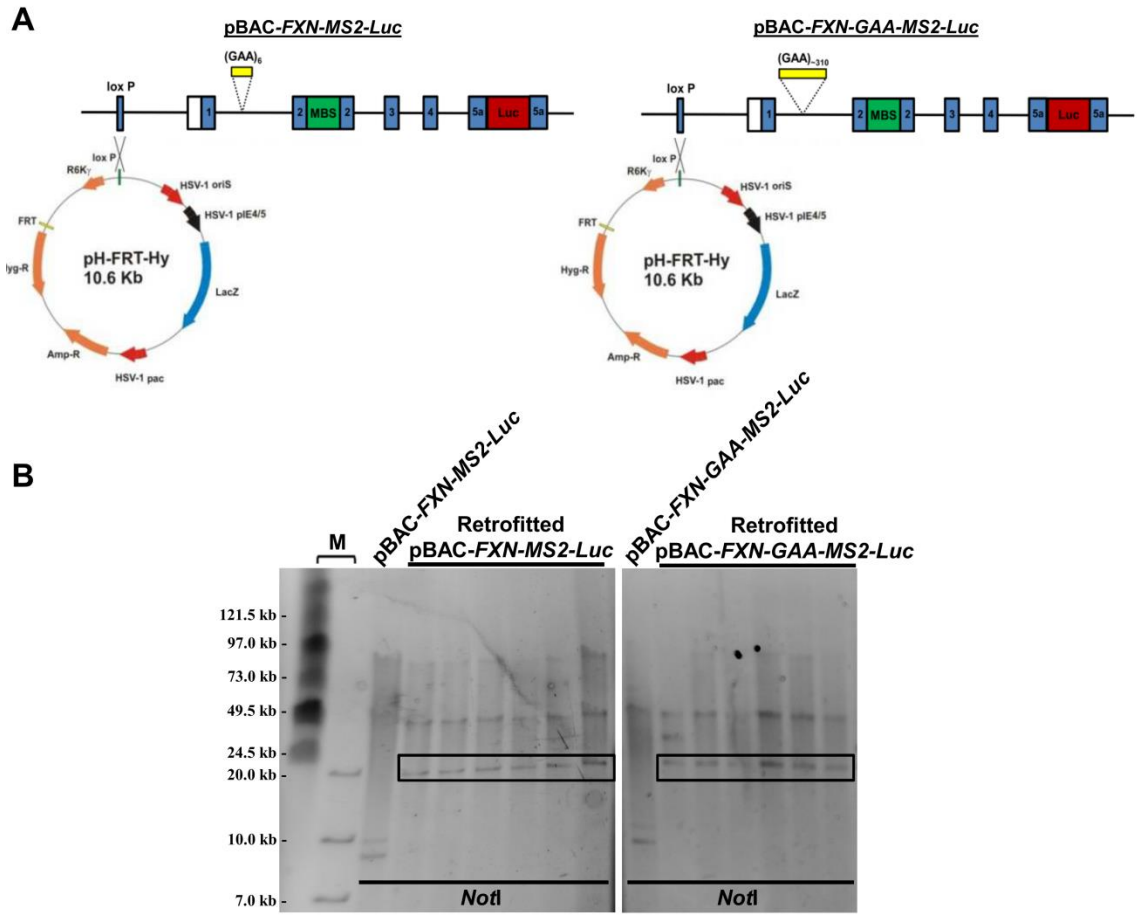

Figure 4.9 Cre-*loxP* mediated retrofitting of pBAC-*FXN-MS2-Luc* and pBAC-*FXN-GAA-MS2-Luc* vectors to pH-FRT-Hy. (**A**) pBAC-*FXN-MS2-Luc* and pBAC-*FXN-GAA-MS2-Luc* were retrofitted with pH-FRT-Hy, which contains a FRT site, a promoter-less hygromycin cassette and a *lacZ* gene, by Cre-mediated recombination between *loxP* sites. (**B**) Integration of a single copy of pH-FRT-Hy produces a 19.35 kb fragment after *Not*I digestion.

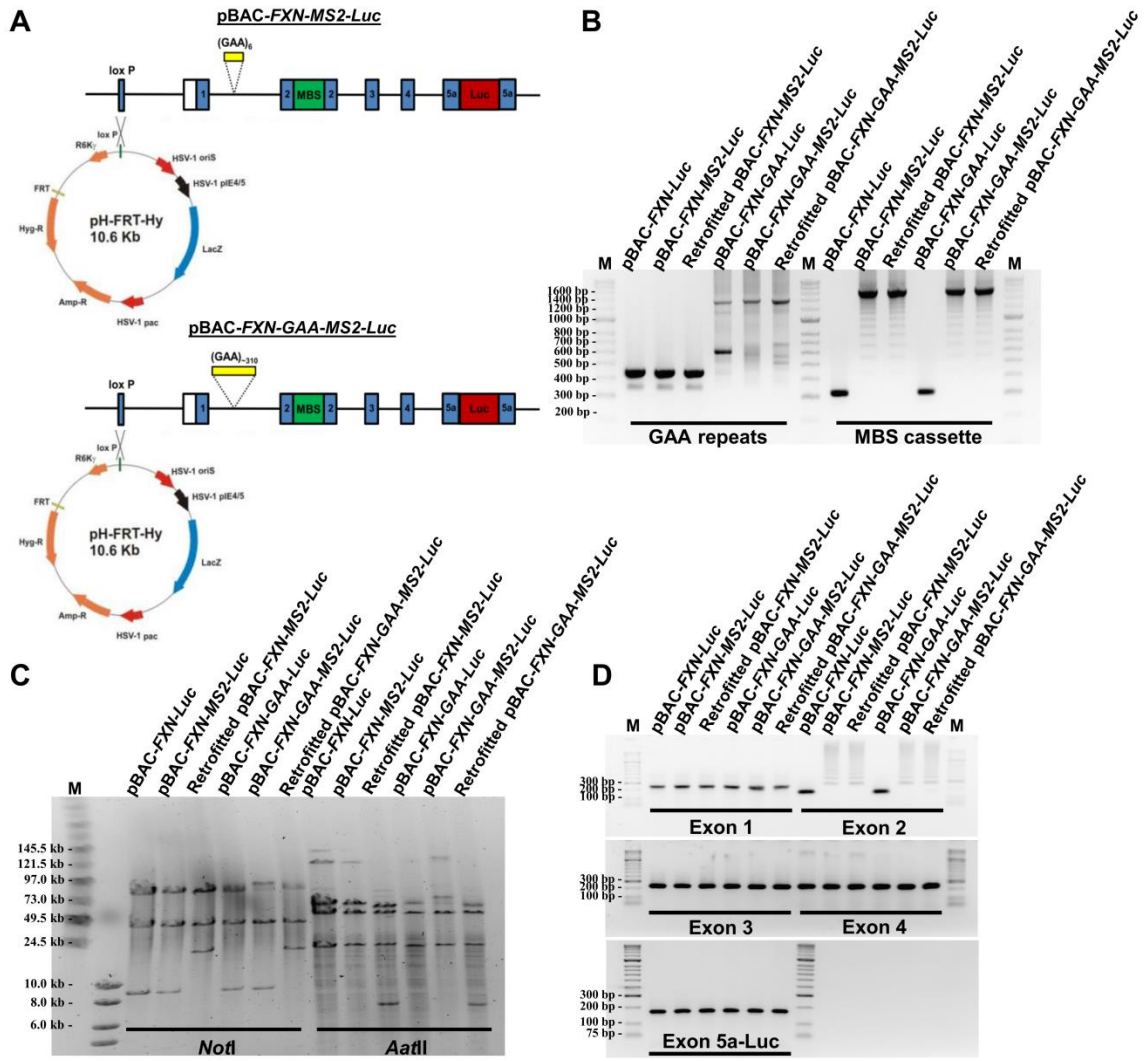

Figure 4.10 Confirmation of successful construction of the pBAC-*FXN-MS2-Luc* and pBAC-*FXN-GAA-MS2-Luc* vectors from pBAC-*FXN-Luc* and pBAC-*FXN-GAA-Luc.* (**A**) Schematic representation of the pBAC-*FXN-MS2-Luc* and pBAC-*FXN-GAA-MS2-Luc* vectors. Each vector carries either six or ~310 GAA repeats in intron 1, respectively, an array of 24 MBS in exon 2 and expresses a FXN-luciferase fusion protein. For stable cell line generation, vectors were retrofitted with pH-FRT-Hy to insert a promoter-less hygromycin cassette. (**B**) Insertion and sizing of GAA repeats and MBS cassette by PCR analysis. (**C**) BAC integrity was confirmed by *Not*I and *Aat*II digestion followed by pulsed-field gel electrophoresis. (**D**) Presence of all exons in the constructs was analysed by PCR.

4.2.2 Transient expression of the *FXN-MS2-Luc* constructs

Quantification of the effect of ~310 GAA repeats on *FXN-GAA-MS2-Luc* expression was achieved by transient expression in HEK 293 cells. Successful transfection of pBAC-*FXN-MS2-Luc* and pBAC-*FXN-GAA-MS2-Luc* was confirmed by LacZ staining (Fig. 4.11A). Expression throughout each transgene was analysed by RT-PCR using primers hybridising with the MBS and Luciferase sequences (Fig. 4.11B). Exogenous transcript was detected in both pBAC-*FXN-MS2-Luc* and pBAC-*FXN-GAA-MS2-Luc* transfected

cells when compared with untransfected HEK cells, demonstrating that the insertion of the MBS cassette does not disrupt transcription of the transgene and confirming the physiological relevance of these constructs. The effect of the ~310 GAA repeat expansion in the pBAC-*FXN-GAA-MS2-Luc* construct was analysed by qRT-PCR and Luciferase assay, yielding a similar transgene repression as obtained with the pBAC-*FXN-GAA-Luc* (Chapter 3) (Lufino *et al*, 2013). *FXN-GAA-MS2-Luc* mRNA levels were reduced by 49% (Fig. 4.11C) and frataxin protein levels by 45% (Fig. 4.11D) when compared with the pBAC-*FXN-MS2-Luc* construct.

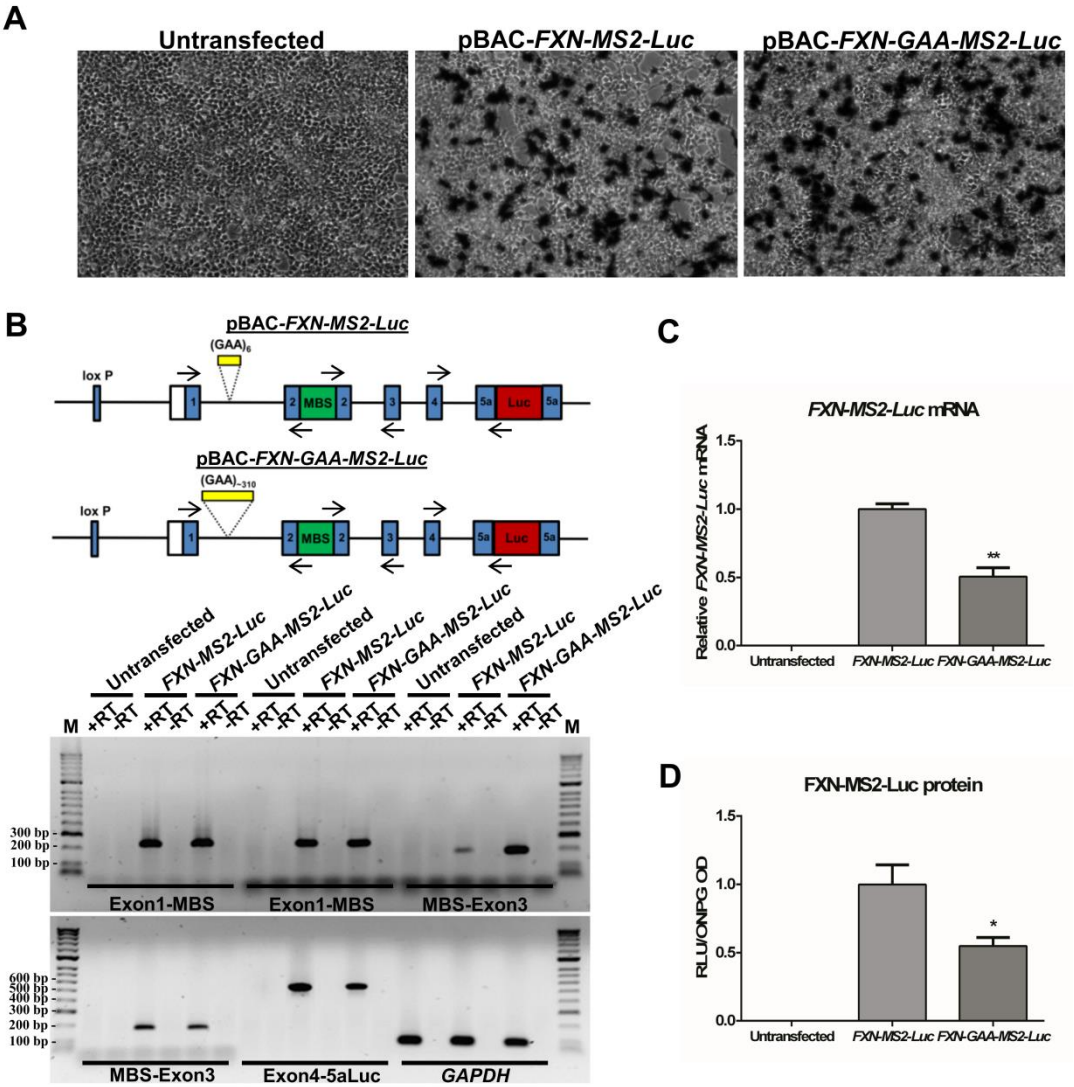

Figure 4.11 Transient expression of the *FXN-MS2-Luc* vectors in HEK293 cells. (**A**) Successful transfection of pBAC-*FXN-MS2-Luc* and pBAC-*FXN-GAA-MS2-Luc* into HEK 293 cells was confirmed by LacZ staining. (**B**) Vector expression was analysed by RT-PCR using two different sets of primers hybridising upstream and downstream of the MBS, and from exon 4 to 5a-luciferase. *GAPDH* mRNA expression was used as control. (**C-D**) Expanded GAA repeats caused a significant reduction in *FXN-GAA-MS2-Luc* mRNA (C) and protein (D) levels, as determined by unpaired two-tailed Student's t-test (**P* < 0.05, ***P* < 0.01). Data are mean ± SEM (*n* = 3).

4.2.3 Imaging optimisation

Single-cell detection of mRNAs synthesised from an MBS-tagged allele can be performed by imaging fixed or living cells. In fixed cells, RNA FISH is performed using a probe that hybridises to the MBS sequence, allowing detection and quantification of single mRNA molecules transcribed from the transgene. In live-cell imaging, nascent mRNA molecules are detected *in vivo* as they are synthesised by the RNAPII and as the MBS is bound by MCP-YFP fusion proteins, allowing time-lapse and transcription kinetics studies (Yunger *et al*, 2010, 2013).

A positive control (pHGCX-Ex2-MS2) was engineered to optimise the RNA FISH and imaging conditions in fixed cells. To generate pHGCX-Ex2-MS2, the PCR product containing the Exon2-MBS sequence – described in the previous section to insert the MBS cassette into pBAC-*FXN-GAA-MS2-Luc* – was inserted in the multiple cloning site of the pHGCX vector. The resulting pHGCX-Ex2-MS2 construct encoded an Exon2-MS2-GFP fusion protein expressed under the control of a CMV promoter (Fig 4.12). HEK 293 cells were transiently transfected with pHGCX-Ex2-MS2 while using the pHGCX plasmid as a negative control. GFP expression in both conditions allowed visual identification of all cells carrying either plasmid using a widefield epifluorescent microscope (Fig. 4.13). Transcription of *CMV-Ex2-MS2-GFP* was analysed by RNA FISH using a probe hybridising to the MBS sequence. Single mRNA molecules were identified as sharp signals above non-specific MS2 fluorescence HEK cell background only in cells containing the pHGCX-Ex2-MS2 plasmid.

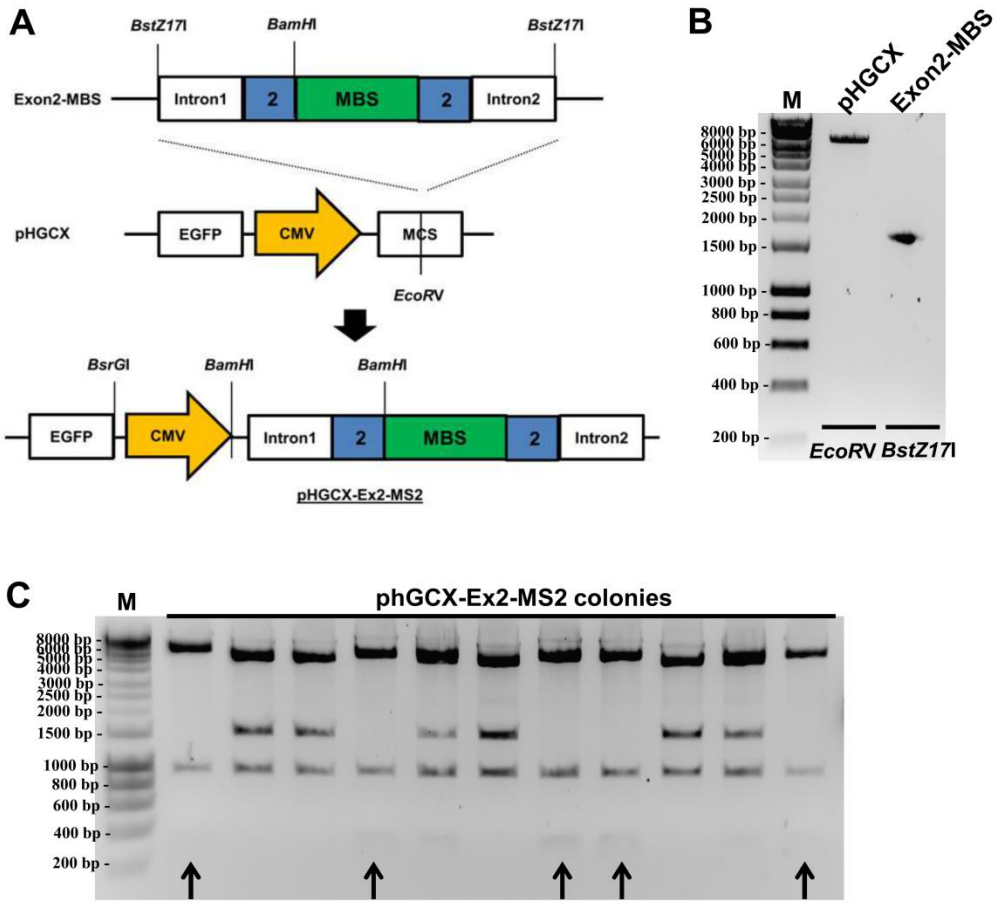

Figure 4.12 Construction of pHGCX-Ex2-MS2 as a positive control for MS2 RNA FISH.

(A) Schematic representation of insertion of Ex2-MBS into the multiple cloning site (MCS) of pHGCX to generate pHGCX-Ex2-MS2, which encodes an Exon2-MS2-GFP fusion protein expressed under the control of a CMV promoter. **(B)** pHGCX was linearised by *EcoR*V digestion and Exon2-MBS was excised from pEx2-MBS by *BstZ17*I digestion. Fragments were ligated and transformed into *E. coli*. **(C)** Successful insertion of the MBS cassette was confirmed by colony-PCR. (**D**) Insertion and correct orientation of Ex2-MBS was confirmed by *BamHI*/*BsrG*I digestion. Arrows indicate insert in the correct orientation.

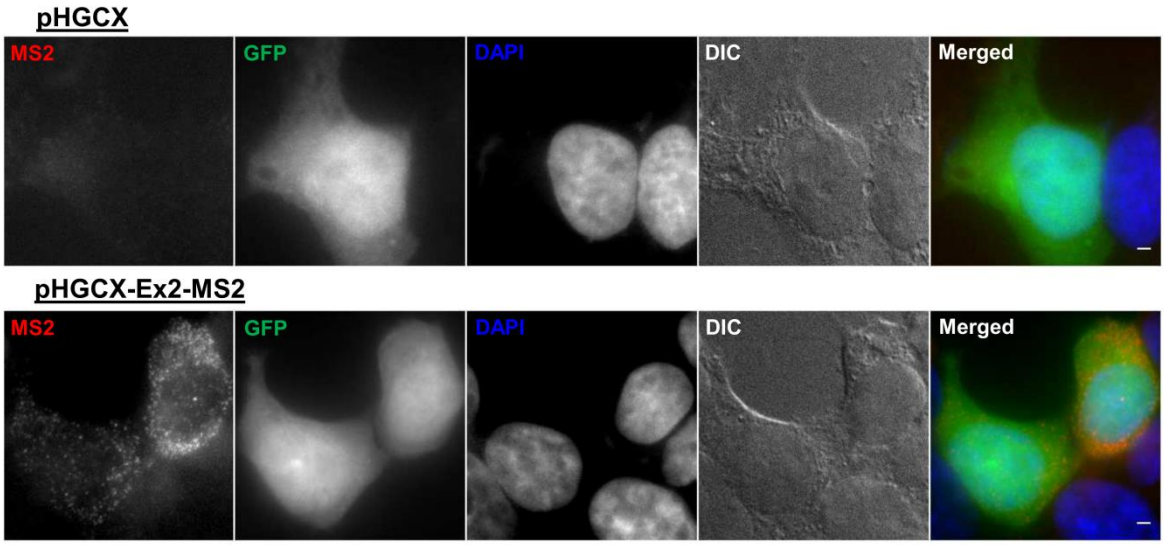

Figure 4.13 MS2 imaging optimisation in fixed HEK 293 cells. Detection of *CMV-Ex2-MS2-GFP* mRNA molecules by RNA FISH. MS2-tagged mRNA molecules were identified as sharp dots above background of non-specific fluorescence using an MS2-Cy3 probe (pseudocoloured red in merged image) in HEK 293 cells transfected with pHGCX-Ex2-MS2. GFP expression (pseudocoloured green) allowed identification of transfected cells. pHGCX was used as a negative control since no MS2-tagged mRNA was expressed. Nuclei were identified by DAPI staining (psudocoloured blue). Images are maximum intensity projections of z-stacks covering the signals. Scale bar 2 µm.

To optimise live-cell imaging conditions, HEK FRT cells were transiently co-transfected with the MCP-YFP and either pBAC-*FXN-MS2-Luc* or pBAC-*FXN-GAA-MS2-Luc*. Two version of the MCP-YFP were used: MCP-YFP-NLS and NES-YFP-MS2-NLS. The original MCP-YFP contains a nuclear localisation signal (NLS) resulting in the nuclear sequestration of unbound fluorescent coat protein molecules in the nucleus, and thus reduction of the cytoplasmic background signal. This facilitates single mRNA detection in the cytoplasm, which is specially suitable for studies of cytoplasmic localisation and dynamics of mRNAs (Bertrand *et al*, 1998; Fusco *et al*, 2003), but may obstruct mRNA detection in the nucleus because the background is too strong. To improve signal-to-noise ratio, the NES-YFP-MS2-NLS (referred throughout the text as MCP-YFP) containing a nuclear export signal (NES) to balance out the NLS was used, resulting in a low background throughout the whole cell and allowing nuclear events to be monitored (Ben-Ari *et al*, 2010). Expression of this improved MCP-YFP protein in cells containing the MBS-tagged RNA led to the fluorescent labelling and detection of active transcription sites in real-time via specific binding between the MCP-YFP protein and the MBS-RNA stem-loops (Fig. 4.14). Within the cell population carrying either pBAC-*FXN-MS2-Luc* or pBAC-*FXN-GAA-MS2-Luc*, some cells presented bright spots in the nucleus that could be

followed over time (Fig. 4.14; movie 1 and 2). Cells expressing only the MCP-YFP protein did not present any nuclear spots.

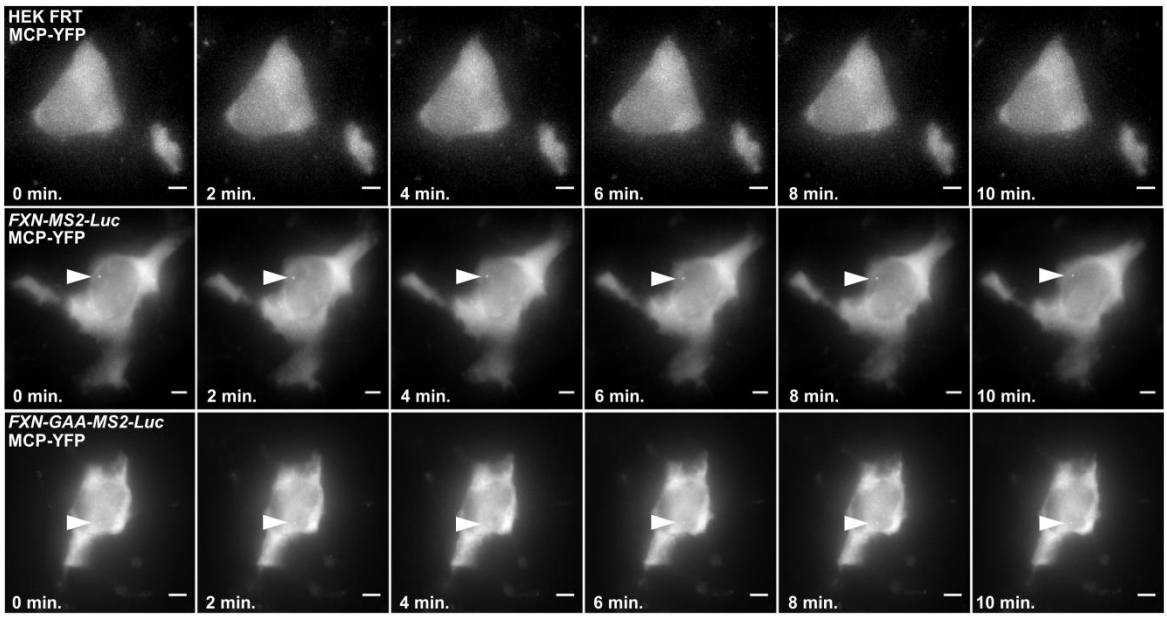

Figure 4.14 Live-cell imaging optimisation in HEK FRT cells. Live-cell fluorescence movie frames taken at indicated times showing nascent *FXN* transcripts in living HEK FRT cells after transient co-transfection of pBAC-*FXN-MS2-Luc* and pBAC-*FXN-GAA-MS2-Luc* vectors with MCP-YFP. HEK FRT cells transfected with only MCP-YFP were a negative control. Cells were captured as 3D stacks every 2 minutes, for 4 hours. Arrows indicate transcription spots. Images are maximum intensity projections of 3D stacks. Scale bar 5 µm.

4.2.4 Generation of stable *FXN-MS2-Luc* cell lines

Several stable human clones of *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* were generated by transfecting HEK FRT cells with either pBAC-*FXN-MS2-Luc* or pBAC-*FXN-GAA-MS2-Luc*, respectively. The Flp-In recombination system uses the Flp recombinase and site-specific recombination to promote the integration of pBAC-*FXN-MS2-Luc* or pBAC-*FXN-GAA-MS2-Luc* into a specific FRT site in chromosome 1 of HEK FRT cells (Fig. 4.15A). These stable cell lines carried a single copy of each vector in the same integration site (Fig. 4.15B-E), allowing a direct single-allele comparison of normal and expanded *FXN* loci. I selected *FXN-MS2-Luc* clone 2 and *FXN-GAA-MS2-Luc* clone 22 (referred to throughout the text as *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc*, respectively) for the following studies, as these best represented the average reduction in *FXN* mRNA (Fig. 4.16A) and protein levels (Fig. 4.16B) in screened clones. The ~310 GAA repeat expansion in the *FXN-GAA-MS2-Luc* cell line recapitulates the characteristic *FXN* gene

repression seen in FRDA by altering the *FXN* epigenetic landscape (Chapter 3) and reducing *FXN* mRNA levels by 62% and frataxin protein levels by 52% (Fig. 4.16).

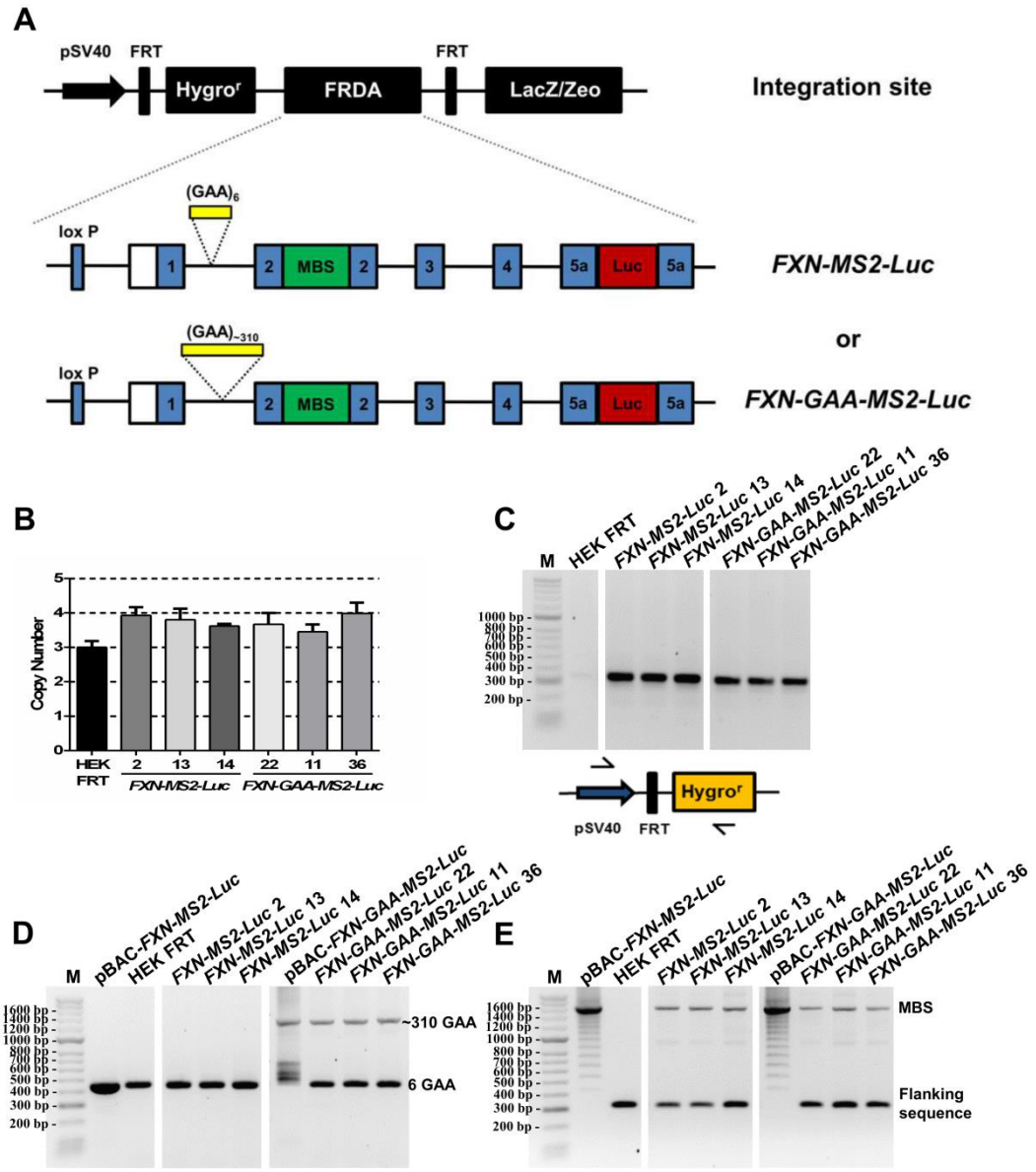

Figure 4.15 *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cell models to study *FXN* localisation and expression. (**A**) Schematic representation of the pBAC-*FXN-MS2-Luc* and pBAC-*FXN-GAA-MS2-Luc* vectors integrated in an FRT site in chromosome 1 of HEK FRT cells. (**B**) Single-copy clones were identified by real-time PCR using UpGAA primers normalised by *GAPDH*. Data are relative to HEK FRT, which contains three endogenous copies. Data are mean ± SEM (n = 3). (**C**) Site-specific vector integration in single-copy *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* clones was confirmed by PCR using primers targeting the P_{SV40} promoter and the hygromycin cassette. (**D-E**) Insertion and sizing of GAA repeats (D) and MBS cassette (E) was determined by PCR analysis in pBAC-*FXN-MS2-Luc*, pBAC-*FXN-GAA-MS2-Luc*, single-copy *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* clones. (**C-E**) Vertically sliced images are juxtaposed lanes that were non-adjacent in the same agarose gel.

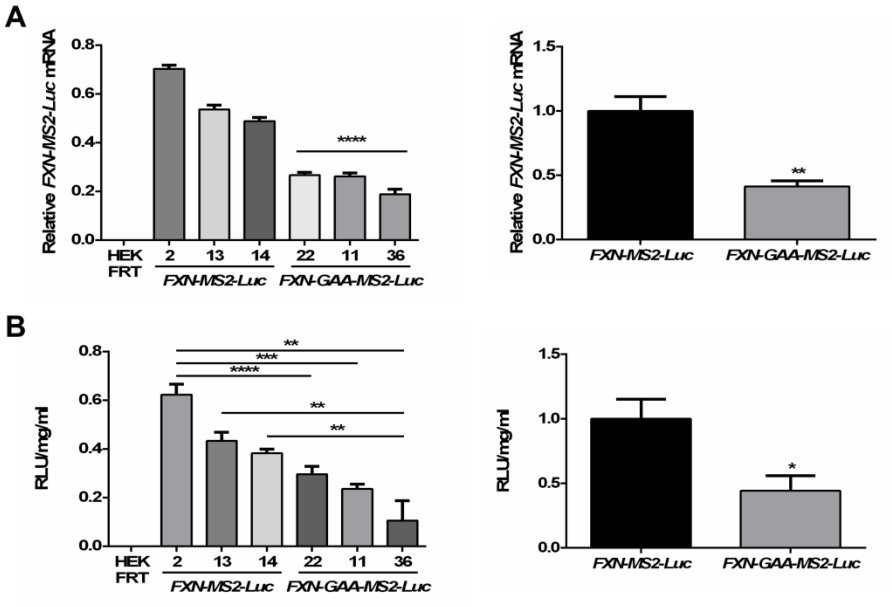

Figure 4.16 Expression of *FXN-MS2-Luc* mRNA and protein in *FXN-MS2-Luc* cell lines. (**A**) *FXN-MS2-Luc* mRNA expression was determined by qRT-PCR using qFXN-Luc primers in single-copy clones (left). One-way analysis of variance (ANOVA) with Tukey's multiple comparisons test showed a significant reduction in *FXN-GAA-MS2-Luc* mRNA levels (****P < 0.0001). Data are mean ± SEM (n = 3). The average of *FXN-MS2-Luc* mRNA expression in all single-copy clones (right) showed a significant reduction in *FXN-GAA-MS2-Luc* mRNA levels, as determined by unpaired Student's t-test (**P < 0.01). Data are mean ± SEM (n = 3). (**B**) Luciferase assay showed a significant reduction in FXN-MS2-luciferase protein levels in *FXN-GAA-MS2-Luc* cells. Data were expressed as relative light units (RLU) and normalised to total cell protein (left). Data are mean ± SEM (n = 3). **P < 0.01, ***P < 0.001, ****P < 0.0001, as determined by One-way ANOVA with Tukey's multiple comparisons test. The average of FXN-MS2-Luciferase protein levels in all single-copy clones (right) showed a significant reduction in *FXN-GAA-MS2-Luc*, as determined by unpaired Student's t-test (*P < 0.05). Data are mean ± SEM (n = 3).

4.2.5 GAA repeat expansion increases *FXN-GAA-MS2-Luc* positioning at the NL

FXN transgene localisation in *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* lines was determined by immunofluorescence *in situ* hybridisation (Immuno-FISH), using co-localised probes for the vector backbone and *FXN* loci to distinguish the transgene from the three additional endogenous *FXN* loci present in each clone (Fig. 4.17A). The *FXN* transgene was found at the NL in ~44% of *FXN-GAA-MS2-Luc* cells compared to only ~10% of *FXN-MS2-Luc* cells. Treatment with the HDAC inhibitors nicotinamide and compound 106 – previously shown to upregulate *FXN* expression and increase histone acetylation levels and DNAse I accessibility at the expanded *FXN* locus in FRDA patient cells (Xu *et al*, 2009; Chan *et al*, 2013) – repositioned only *FXN-GAA-MS2-Luc* away from the NL (Fig. 4.17B), but increased transgene *FXN* mRNA levels in both cell lines (Fig.

4.17C), suggesting a complex interplay between *FXN* repression and intranuclear localisation.

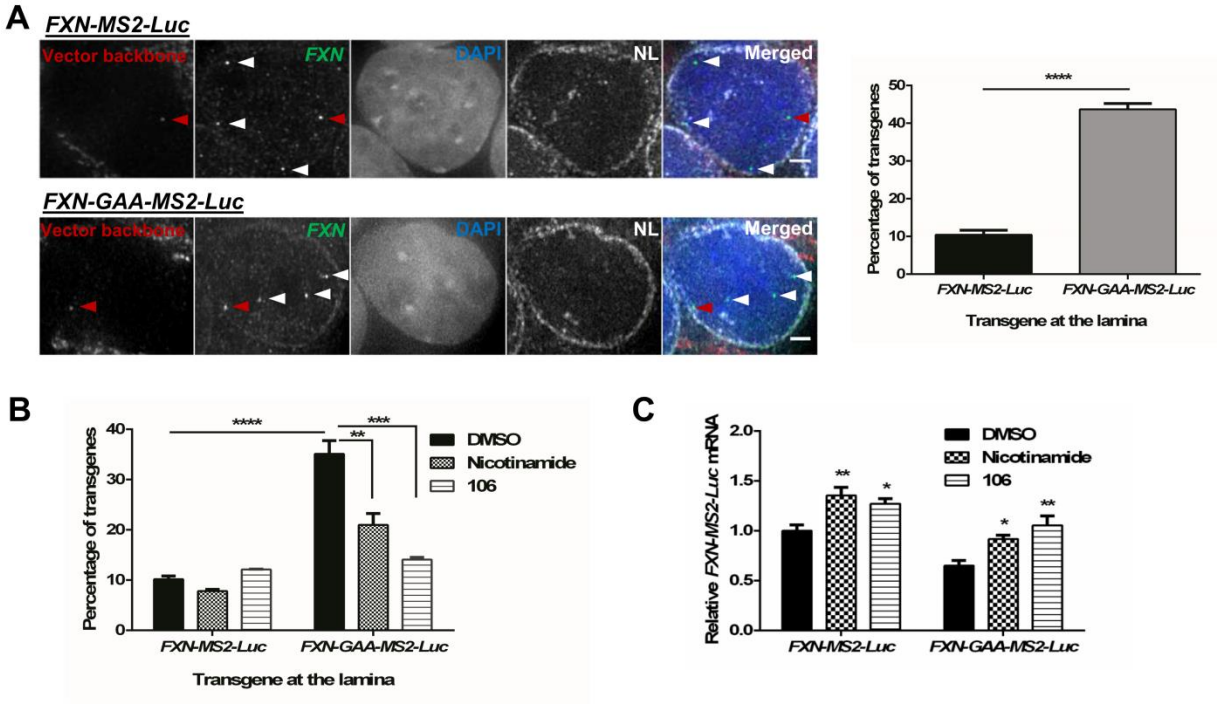

Figure 4.17 The expanded GAA repeat *FXN* transgene associates with the NL more frequently in an *FXN-GAA-MS2-Luc* cell model. (**A**) 3D Immuno-FISH detection of *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* association with the NL (pseudocoloured white in merged image). Transgenes were identified by co-localisation of FISH signals from *FXN* loci (green) and the vector backbone (red) in *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells (left). Nuclei were identified by DAPI staining (blue). Arrows show endogenous (white) and transgene (red) alleles. Comparison of the proportion of transgenes contacting the NL showed a significant difference in *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells (right) (*****P* < 0.0001, as determined by Fisher's two-tailed exact test, *n* = 140 cells per condition). Data are mean ± SEM from two independent FISH experiments. Association of FISH signals with the NL was analysed using ImageJ: image stacks were analysed manually and association was determined by eye when *FXN* FISH signals were touching or superimposing the NL signal at a given z-section. Images are maximum projections of z-stacks covering the signals. Scale bar 2 µm. (**B**) Treatment with nicotinamide (10 mM) or compound 106 (10 µM) for 48 hours significantly decreased *FXN-GAA-MS2-Luc* interaction with NL (***P* < 0.01, ****P* < 0.001, *****P* < 0.0001, as determined by Fisher's two-tailed exact test, *n* = 140 cells per condition). Data are mean ± SEM from 2-3 independent FISH experiments. (**C**) *FXN-MS2-Luc* mRNA expression after incubation with HDACi for 48 hours was determined by qRT-PCR using qFXN-Luc primers. Data were normalised to *GAPDH* mRNA levels. Results are relative to *FXN-MS2-Luc* DMSO. Two-way ANOVA with Sidak's correction test showed a significant difference in the two cell lines when comparing with the DMSO control of each line (**P* < 0.05, ***P* < 0.01). Data are mean ± SEM from three independent experiments.

4.2.6 GAA-expanded repeats decrease the number of *FXN* mRNA molecules and slow transcription kinetics in single *FXN-GAA-MS2* cells

To further understand *FXN* repression, the transcriptional output of individual transgenic *FXN* alleles in *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells was analysed by RNA FISH using a probe hybridising to the MBS sequence as optimised in section 4.2.3. Single *FXN-MS2-Luc* mRNA molecules were identified as sharp signals above non-specific fluorescence HEK cell background in accordance with previous studies demonstrating single-molecule resolution of this technique (Fig. 4.18A and C) (Fusco *et al*, 2003; Raj *et al*, 2008; Yunger *et al*, 2010). The insertion of the MBS in exon 2 led to the identification of all fluorescence spots as mature *FXN* mRNA molecules, considering transcription sites were not distinguishable. No difference was detected in the percentage of cells expressing the transgene (Fig. 4.18B). *FXN-GAA-MS2-Luc* cells contained ~5 ± 2 mRNAs per cell, and *FXN-MS2-Luc* cells contained ~9 ± 4 mRNA per cell (Fig. 4.18C), therefore ~310 GAA repeats reduce the number of mature mRNA molecules by 44% at single-cell level. This quantification is in agreement with oligonucleotide array assays previously showing the majority of mRNAs are present at <10 copies per cell (Lockhart & Winzeler, 2000).

I then followed the kinetic behaviour of *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* transcription using four-dimensional time-lapse imaging and fluorescence recovery after photobleaching (FRAP) (Yunger *et al*, 2010). Nascent transcripts were detected as fluorescent spots as they were bound by MCP-YFP fusion protein and visualised in real-time. Active transcription sites were detected in both cell lines over long periods of time using a widefield fluorescent microscope, albeit in only a subset of cells. Given the constant expression of *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc*, no pulsing activity was detected (Fig. 4.19; Movie 3 and 4).

Recovery of fluorescence indicates new transcripts are actively synthesised at the transcription site (Fig. 4.20A; Movie 5 and 6). FRAP curves showed that the time required to reach steady state after photobleaching was different between the two cell lines. *FXN-MS2-Luc* cells showed half-time of recovery ($t_{1/2}$) = 12.90 s with full recovery = 120 s (Fig. 4.20B and Movie 5), whereas *FXN-GAA-MS2-Luc* cells showed slower kinetics $t_{1/2}$ = 48.10 s with full recovery = 260 s (Fig. 4.20B; Movie 6). Since I found no difference in *FXN* mRNA stability in *FXN-GAA-MS2-Luc* and *FXN-GAA-MS2-Luc* cells (Fig. 4.20C) (Punga & Bühler, 2010), RNA FISH and FRAP data indicate that the GAA expansion decreases the amount of *FXN-GAA-MS2-Luc* mRNA molecules by impeding RNAPII transcription

initiation and/or elongation. However, I could not distinguish between the two events given the placement of the MBS in exon 2, where the FRAP readout is obtained.

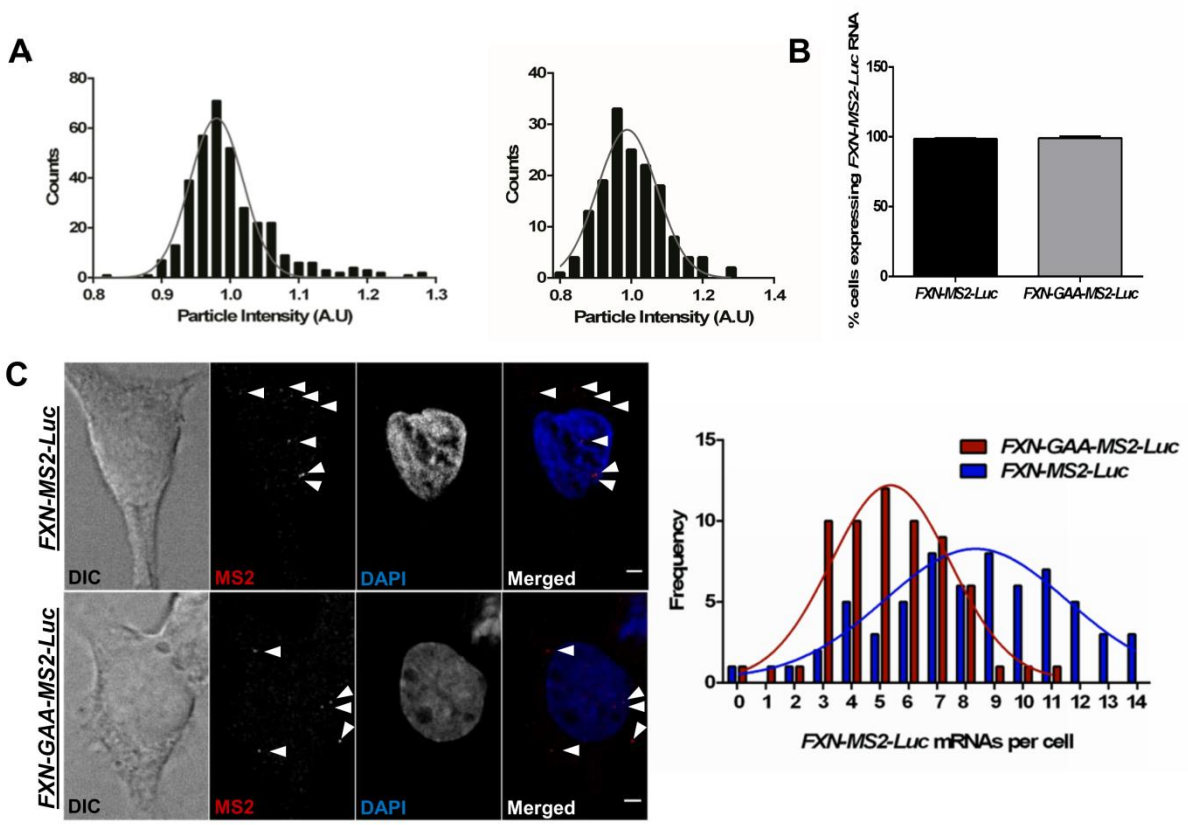

Figure 4.18 The GAA repeat expansion reduces *FXN-GAA-MS2-Luc* transcriptional output in fixed *FXN-GAA-Luc* cells. (**A**) RNA FISH spot intensity histogram of *FXN-MS2-Luc* (left) and *FXN-GAA-MS2-Luc* (right) showing all measurements resulting in a Gaussian distribution with a mean that represents the most likely fluorescence signal of a single molecule. The lower tail of this distribution represents noise due to aggregates of the MS2 FISH probe and the high-value tail represents mRNA aggregations. RNA particles contained only single copies of *FXN-MS2-Luc* RNA in *FXN-MS2-Luc* cell lines (n = 34). Curves show a one-component Gaussian fit. (**B**) Proportion of *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells (n = 63 cells per line) with ≥1 RNA FISH signals were identical. Data are mean ± SEM from two independent FISH experiments. (**C**) Quantification of *FXN-MS2-Luc* mRNA molecules in *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells by RNA FISH. *FXN-MS2-Luc* mRNA molecules were identified as sharp dots (arrows) above background of non-specific fluorescence using an MS2-Cy3 probe (red in merged images) in *FXN-MS2-Luc* (top images) and *FXN-GAA-MS2-Luc* (bottom images) cells. Non-linear regression curve fit comparison showed a significant difference between the distribution of *FXN-MS2-Luc* mRNA molecules in *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells (P < 0.0001, (F = 33.31, d.f. = 3, 21), n = 63 cells per line) (right). Images are maximum projections of Z stacks covering the signals. Scale bar 2 µm.

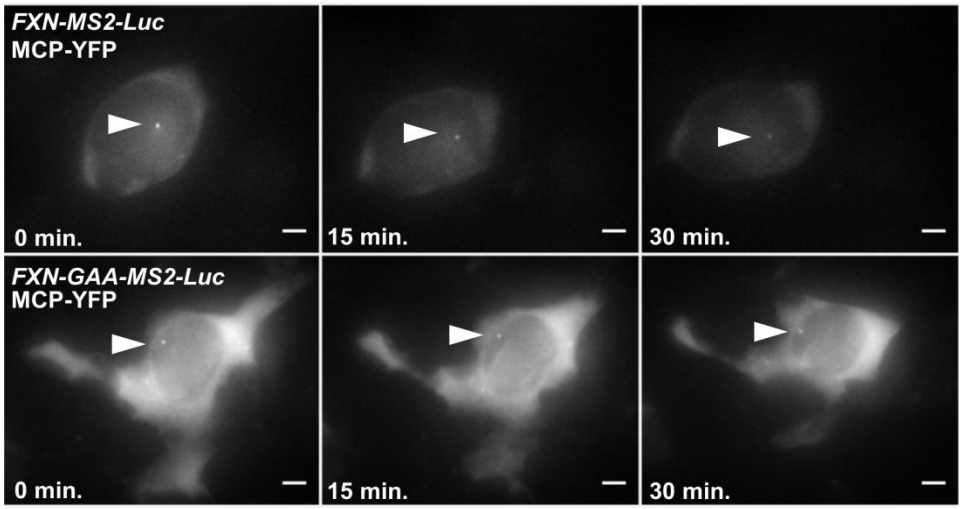

Figure 4.19 Real time visualisation of *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* transcription. Nascent *FXN-MS2-Luc* (top) and *FXN-GAA-MS2-Luc* (bottom) mRNA were detected as fluorescent spots as transcripts were bound by MS2-YFP fusion protein. Exposure: 30 ms; cells were captured as 3D stacks every 2.5 minutes, for 50 minutes. The displayed images are maximum-projected z-stacks taken at indicated times. Scale bar 5 µm.

To elucidate the link between *FXN* localisation and repression, I analysed the fluorescence intensities of active transgenes in living *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells. The fluorescence intensity of transcription sites was significantly lower when transgenes were expressed at the nuclear periphery (NP) compared to the interior in both *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells (Fig. 4.20D). This indicates that both *FXN* transgenes express when localised at the NP, although they do so at lower levels. When compared to active *FXN-MS2-Luc* transgenes, spot intensity of *FXN-MS2-GAA-Luc* transgenes was significantly lower only in the nuclear interior. Taken together with data from Fig. 4.17A, these data suggest that expanded GAA repeats increase *FXN-GAA-MS2-Luc* positioning at the NL, where expression levels are reduced when compared to the nuclear interior or with *FXN-MS2-Luc* expression levels in the nucleoplasm.

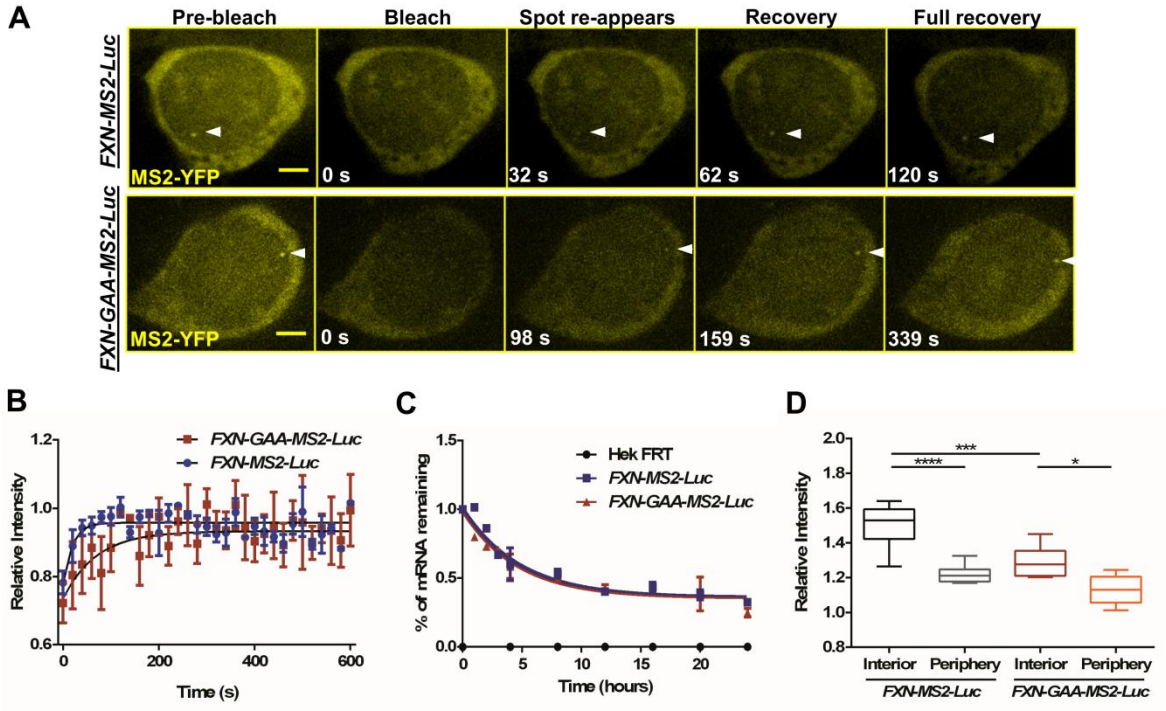

Figure 4.20 The GAA repeat expansion reduces *FXN-GAA-MS2-Luc* transcriptional output and impedes transcription in living *FXN-GAA-Luc* cells. (**A**) Live-cell fluorescence movie frames taken before and at indicated times after photobleaching of the transcription site. Scale bar 2 µm. (**B**) FRAP recovery curves and fitted exponential curves show a difference in transcription kinetics ($P = 0.0048$, ($F = 4.434$, d.f. = 3, 210), $n = 5$ cells per line). (**C**) Degradation of *FXN-MS2-Luc* mRNA was analysed by qRT-PCR using qFXN-Luc primers in *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells after treatment with Actinomycin D for 24 hours. Data were normalised to *β-actin* mRNA. Comparison of fitted exponential decay curves shows no difference between the two cell lines ($P = 0.5979$, ($F = 0.6351$, d.f. = 3, 32). *FXN-MS2-Luc* mRNA half-life was ~4h. Data are mean ± SEM ($n = 3$). (**D**) Quantification of fluorescence intensities of active transgenes in living *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells. Spot intensity of active *FXN* transgenes above MS2-YFP background was significantly lower in the nuclear interior in *FXN-GAA-MS2-Luc* cells when compared to *FXN-MS2-Luc* cells (***$P < 0.001$), and significantly lower when alleles were expressing at the periphery than in the nuclear interior in the two cell lines, as determined by one-way ANOVA with Sidak's multiple comparisons test (****$P < 0.0001$, *$P < 0.05$, $n = 15$ cells per line).

4.3 Discussion

NL-associated chromatin at the NP is mostly transcriptionally inactive and enriched in repressive histone modifications, such as H3K9me2 and H3K27me3 (Kind & van Steensel, 2010). Although it is still unclear if gene positioning affects transcription, chromatin remodelling is sufficient to induce nuclear reorganisation (Therizols *et al*, 2014). It is generally assumed that GAA expansion–mediated epigenetic changes in *FXN* locus can reduce *FXN* mRNA levels in averaged cell samples though a poorly understood mechanism. Working for the first time at the single-cell level, the work described in this

Chapter identifies the NL as a novel player in *FXN* transcriptional impairment. I show that expanded GAA repeats increase *FXN* positioning at the NL, where HDACs can deacetylase histones and help repress gene activity. A marked increase in *FXN-GAA-MS2-Luc* positioning at the NL with a proportional reduction in the number of *FXN-GAA-MS2-Luc* mRNA molecules suggests that expanded transgene localisation at the NL leads to *FXN* impairment and silencing. The mechanism of retention at the NL of these expanded repeat structures is highly sensitive to HDAC inhibition. *FXN* transcription from both normal and expanded transgenes is upregulated by restoring acetylation by similar amounts, yet results in significant movement from the NL of only the expanded alleles. However, the expanded alleles are unable to achieve the same level of upregulation as their normal counterparts. This could be the result of a different extent of histone acetylation achieved by normal and expanded alleles following treatment with HDAC inhibitors, which could be caused by reduced accessibility of compacted *FXN* alleles. However, the observed difference in *FXN* upregulation could also imply that other structural changes such as R-loop formation contribute to transcriptional dysregulation and that the trigger event is induced by structural and epigenetic changes to the DNA followed by an increased likelihood of retention at the NL.

Using the MS2 system, I show that active *FXN-GAA-MS2-Luc* alleles present impaired transcription kinetics in living single cells. It would be possible to further dissect the *FXN* impairment by using the dual-labelling system with MS2 and PP7 cassettes in different regions of the expanded *FXN* locus. Alternatively, isogenic cell clones in which the *FXN-GAA-MS2* transgene is always integrated into the same genomic FRT site, but containing the MS2 cassette in different locations throughout the gene could be generated to compare different RNAPII transcription rates.

Variability in gene expression has been linked to dynamic changes in the state of chromatin taking place locally at the gene locus, leading to a probabilistic promoter switch between active and inactive state (Coulon *et al*, 2013). Time-lapse experiments did not reveal bursting activity either in the normal or expanded transgene. Given that *FXN* expression is essential for life (Cossée *et al*, 2000), it could be speculated that the constitutive expression of the *FXN* locus observed here may resemble that of housekeeping genes, where polymerases fire individually (Coulon *et al*, 2013).

Further studies in healthy and FRDA patient-derived cells discussed in Chapter 5, will elucidate the link between expanded *FXN* positioning at the NL and GAA-mediated transcriptional repression.

CHAPTER 5

FXN gene nuclear localisation and expression in healthy, carrier and FRDA patient-derived cells

The results featured in this Chapter were published in the paper (Annex 2):

Silva AM, Brown JM, Buckle VJ, Wade-Martins R and Lufino MM (2015) Expanded GAA Repeats Impair *FXN* Gene Expression and Reposition the *FXN* Locus to the Nuclear Lamina in Single Cells. *Hum. Mol. Genet.* 24(12):3457-71.

Author contributions:

A.M.S, J.M.B., V.J.B., R.W.-M. and M.M.L. conceived and designed the experiments; **A.M.S** performed the experiments and analysed the data; **A.M.S** wrote the manuscript; **A.M.S**, J.M.B., V.J.B., R.W.-M. and M.M.L. revised the manuscript.

5.1 Introduction

Expanded DNA repeats can have a broader impact on nuclear structure and induce an abnormal association of expanded genes or mRNA with certain nuclear compartments. For example, in DM1, patient-derived fibroblasts exhibit altered localisation of the nuclear envelope proteins emerin and lamins A/C and B1 with concomitant increased size and altered shape of nuclei (Rodríguez *et al*, 2015). Moreover, there is evidence of nucleoli distortion in DM1 fibroblasts (Rodríguez *et al*, 2015). Another study reports that CUG-expanded *DMPK* mRNA foci accumulate at the periphery of nuclear speckles, blocking a stage in the mRNA export pathway (Holt *et al*, 2007). In *C9ORF72*-related FTD/ALS, patient cells present nucleolar stress due to the binding of nucleolin, an essential protein of the nucleolus, to the GGGGCC repeat-containg RNA transcripts (Haeusler *et al*, 2014).

During recent years, the epigenetic changes occurring at the expanded *FXN* locus, such as increased DNA methylation at specific CpG sites (Greene *et al*, 2007; Al-Mahdawi *et al*, 2008; Castaldo *et al*, 2008; Evans-Galea *et al*, 2012), reduced histone acetylation (H3/H4ac) and increased levels of methylated histones H3K9me2 and H3K9me3 (Herman *et al*, 2006; Al-Mahdawi *et al*, 2008), have become the target of potential therapies for FRDA. In this Chapter, I aim to uncover a higher level of regulation of *FXN* repression by analysing the interactions of the expanded *FXN* locus with nuclear components required for expression activation or silencing within the nuclear organisation.

5.1.1 Nuclear organisation

The eukaryotic nucleus is a membrane-bound organelle that separates gene transcription from protein translation in the cytoplasm. Within the interphase nucleus, chromosomes occupy discrete and non-random territories. In flat-ellipsoidal nuclei, e.g. fibroblasts, there is a correlation between the location of chromosome territories (CTs) and the size of chromosomes, with larger chromosomes positioned towards the nuclear interior and smaller chromosomes sitting close to the NP (Bolzer *et al*, 2005). However, gene content seems to play a major role in CT determination. In spherical nuclei, e.g. lymphoblasts, gene-rich chromosomes occupy a more central position, whereas gene-poor chromosomes are at the NP (Boyle *et al*, 2001).

Studies indicate that the radial positioning of a gene within the nucleus correlates with its transcriptional output, but whether a gene is transcribed due to its position, or its position is determined by its transcriptional state, is the subject of current research (Dillon, 2008; Takizawa *et al*, 2008b; Ferrai *et al*, 2010a; Therizols *et al*, 2014). Gene expression

usually takes place in the nucleoplasm where distinct, but highly dynamic specialised nuclear compartments exist (reviewed in Spector, 2006; Zhao *et al*, 2009; Qureshi & Mehler, 2010). These heterogeneous structures can dynamically assemble and disassemble, associate with specific genomic loci, interact with chromatin, move, and undergo other changes to support cellular processes. Live-cell imaging shows that chromatin movements are generally constrained around the nucleoli and NP and more mobile in the nucleoplasm (Chubb *et al*, 2002), where chromosome intermingling occurs via extended chromatin loops (Branco & Pombo, 2006) and genes can interact with various nuclear bodies, including but not limited to transcription factories, nuclear speckles, paraspeckles, Cajal bodies, promyelocytic leukemia (PML) nuclear bodies and nucleoli (Fig. 5.1).

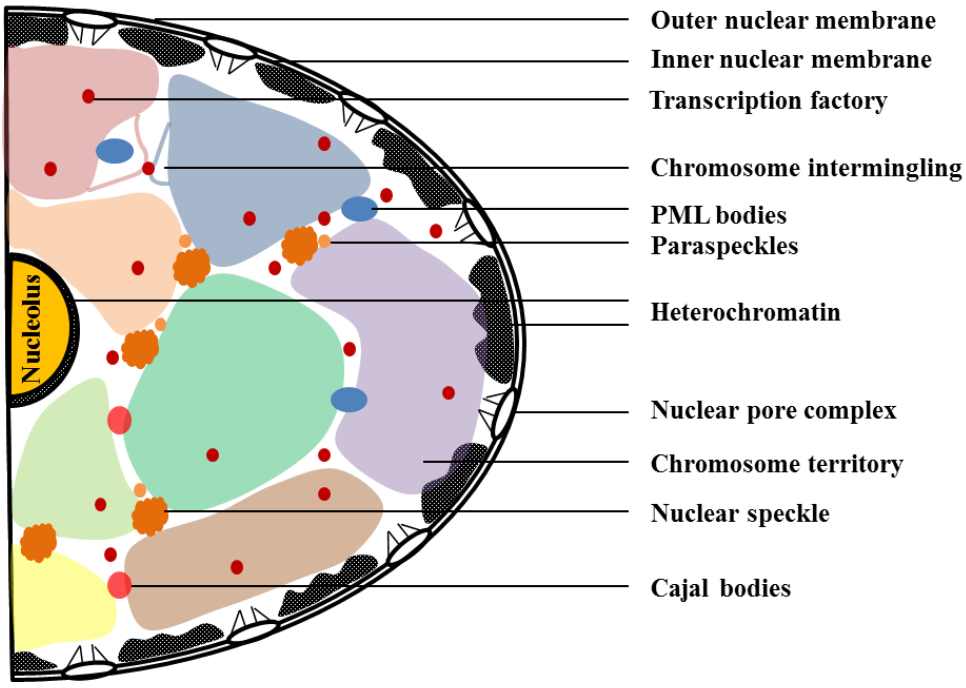

Figure 5.1 Nuclear organisation. Illustration of the mammalian nucleus showing the spatial organisation of nuclear compartments.

5.1.1.1 Nuclear compartments involved in gene expression

Multiple transcription factories can be observed within the nucleus and are composed of many active RNA polymerases and associated transcription factors where nascent RNA is synthesised (Iborra *et al*, 1996). Active genes were shown to be dynamically relocated and simultaneously transcribed by a shared transcription factory (Osborne *et al*, 2004). Two types of transcription factories have been described: "poised transcription factories" and "active transcription factories" depending on the phosphorylation state of RNAPII,

leading to the idea that association of genes with poised factories can rapidly induce gene activation in response to specific stimuli (Ferrai *et al*, 2010b).

Nuclear speckles act as storage/assembly/modification compartments that can supply splicing factors to active transcription sites. These highly dynamic structures are enriched in components of the pre-mRNA splicing machinery, including small nuclear ribonucleoprotein particles (snRNPs), spliceosome subunits, and other non-snRNP protein splicing factors. Speckles can be detected throughout the nucleoplasm as 20-50 irregular shapes varying in size close to active transcription sites (reviewed in Spector & Lamond, 2011). Active genes have been found surrounding nuclear speckles (Shopland *et al*, 2003; Brown *et al*, 2006). For example, the active allele of the monoallelically expressed glial fibrillary acidic protein (*GFAP*) gene was shown to associate with speckles in 70% of cells, while the inactive allele did not (Takizawa *et al*, 2008a). Furthermore, clustering of coordinately expressed erythroid genes can be found at the same nuclear speckle (Shopland *et al*, 2003; Brown *et al*, 2008). Although there is a tight correlation between expressing genes and nuclear speckles, this association is not mandatory for transcription and pre-mRNA splicing.

Paraspeckles are discrete ribonucleoprotein bodies implicated in the regulation of mRNA nuclear export. These structures form around the long non-coding RNA NEAT1 and contain the protein complexes PSF/SFPQ, P54NRB/NONO and PSPC1 in a transcription-dependent manner (Fox *et al*, 2002). Although not actively involved in transcription, paraspeckles may form in association with active genes and coordinate gene expression by ensuring the nuclear retention of certain mRNAs (reviewed in Bond & Fox, 2009). For example, the hyperedited CTN-RNAs escape nuclear export and are efficiently retained in the nucleus, where the majority localise to the paraspeckles. Under stress conditions, CTN-RNA is post-transcriptionally cleaved to produce protein-coding mCAT2 mRNA and rapidly exported into the cytoplasm (Prasanth *et al*, 2005).

Cajal bodies are involved in pre-mRNA splicing, pre-rRNA processing, histone pre-mRNA 3′ end processing, and telomere maintenance implicating them in the biogenesis of RNA-processing complexes. They have also been implicated in the biogenesis of several classes of spliceosomal snRNPs and are enriched in RNA polymerase subunits, specific basal transcription factors, RNA processing factors, and p80/coilin protein. Their number and size vary throughout the cell cycle and are responsive to cell stress and levels of gene transcription (reviewed in Qureshi & Mehler, 2010; Dundr, 2012).

PML bodies are heterogeneous macromolecular protein complexes composed of PML protein and a variety of other proteins, including Sp100, SUMO-1, Daxx, pRB, p53, and BLM, suggesting a role in the regulation of many cellular processes, such as control of cell proliferation, DNA repair, apoptosis, and viral infection. These structures have also been shown to associate with gene-dense, transcriptionally active chromatin regions, including the major histocompatibility complex class I gene cluster region and the *p53* gene locus, for which PML bodies were proposed to modulate chromatin architecture and transcription (reviewed in Bernardi & Pandolfi, 2007). A stably integrated gene locus was shown to interaction with PML bodies in living cells, but ongoing transcription appeared independent of PML body association (Tsukamoto *et al*, 2000). Furthermore, analysis of multiple genomic regions showed that PML bodies associated with genomic regions with high transcriptional activity, but did not directly regulate transcription of these genes (Wang *et al*, 2004). The exact molecular mechanism leading to PML-mediated transcriptional regulation needs further elucidation.

The nucleolus forms around chromosomal loci of tandemly repeated rDNA gene clusters (called the NORs – nucleolus organiser regions) and is the site of rRNA transcription, processing, and ribosome subunit export (McStay & Grummt, 2008). Active NORs present an open chromatin conformation and are transcribe by RNAPI. On inactive NORs, rDNA appears in a heterochromatic state indistinguishable from the surrounding heterochromatin (called perinucleolar heterochromatin), and lack associated RNAPI and RNAPI–specific factors.

5.1.1.2 Nuclear compartment involved in gene repression

Heterochromatin containing repressed genes is usually found surrounding the nucleolus and the NP.

The nucleolus was discussed in the previous section as a transcription factory involved in ribosomal biogenesis. However, the perinucleolar heterochromatin surrounding the nucleolus is composed of satellite DNA that surrounds NORs, and silent rDNA clusters. Furthermore, there is evidence that the perinucleolar heterochromatin can establish and maintain silencing of non-rDNA-related genomic regions (McStay & Grummt, 2008). For example, in 80-90% of nuclei of synchronised female mouse fibroblasts, following X chromosome pairing, the inactive X chromosome is targeted to the perinucleolar region and establishes *Xist* ncRNA-dependent contacts with the nucleolus (Zhang *et al*, 2007).

Genomic DNA interactions with the NP can actively contribute to gene repression (Finlan *et al*, 2008; Reddy *et al*, 2008; Zullo *et al*, 2012). However this is not a general phenomenon (Finlan *et al*, 2008; Kumaran & Spector, 2008), rather is gene-specific, and may depend on multiple parameters such as transcription factor accessibility, promoter strength, existence of insulator elements and pre-existing chromatin marks, which may counteract the mechanisms underlying transcription repression. The NL tends to be in contact with heterochromatin and is associated with markers of gene repression, such as enrichment in histone modifications H3K9me2 and H3K27me3 and depletion of activating histone marks and RNAPII occupancy (reviewed in Kind & van Steensel, 2010). This nuclear compartment is further discussed in section 5.1.2.

5.1.2 Gene regulation at the NP

Breakthrough studies suggested that the intranuclear position of a gene could affect its expression after measuring the radial position of a locus and analysing the association of low gene-density and late-replicating regions with the NP (Croft *et al*, 1999; Boyle *et al*, 2001; Chubb *et al*, 2002; Bolzer *et al*, 2005).

Different components of the NP have distinct roles in transcriptional regulation. The nucleus and cytoplasm communicate via channels in the nuclear envelope, called nuclear pore complexes, which promote bidirectional exchange of proteins, RNAs and ribonucleoproteins. In animal cells, the inner nuclear membrane (INM) is structurally supported by the NL, a meshwork of intermediate filament proteins composed of lamins and lamin-associated proteins. Localisation to nuclear pores can be correlated with decondensed chromatin and transcriptional activation of genes, while the NL is associated with heterochromatin and linked to the repression of genes and repetitive sequences (Kind & van Steensel, 2010; Padeken & Heun, 2014).

Several studies support that the NP, and in particular the NL, represents a repressive environment for gene expression. Most genes whose activity is linked to differentiation and development transition from a peripheral nuclear position to the nuclear interior upon their activation. Examples include *IgH* and *β-globin* during B cell and erythroid cell differentiation, respectively (Kosak *et al*, 2002; Ragoczy *et al*, 2006), and genes of the *hoxB* cluster during development (Chambeyron & Bickmore, 2004). By analysing the two alleles of the monoallelically expressed *GFAP* gene within the same nucleus, Takizawa and colleagues showed that the active allele is generally found more internally when compared with its inactive counterpart (Takizawa *et al*, 2008a).

However, studies artificially targeting genomic regions to the NP showed conflicting results. In one approach, tethering of a reporter gene to emerin, a protein of the INM, resulted in its transcriptional repression (Reddy *et al*, 2008). Finlan and colleagues tethered human chromosomes to LAP2β, another protein of the INM, and showed that this could reversibly repress the expression of some endogenous genes, but not others (Finlan *et al*, 2008). Finally, in a third approach, a genetic locus was targeted to the NL in living mammalian cells, allowing visualisation of transcription in real-time at the NP upon transcriptional induction (Kumaran & Spector, 2008). Varying effects may be due to experimental differences, but may also reflect different microenvironments and chromatin heterogeneity at the NP. However, these studies indicate that, although the NP is generally considered a transcriptionally repressive environment, microdomains of the INM/NL that are transcriptionally permissive may exist (Zhao *et al*, 2009).

5.1.2.1 The NL as a repressive compartment

Mammalian cells express four main lamins: A and C, which are splicing variants encoded by the *LMNA* gene; and B1 and B2 encoded by the *LMNB1* and *LMNB2* genes, respectively. Lamin filaments are important for the assembly, structure, shape, and mechanical stability of the nucleus. They also play a role in the regulation of chromatin organisation and gene expression. Several transmembrane proteins are also part of the NL. INM proteins containing the LEM domain (Lap2β, Emerin and MAN1) interact with a range of transcriptional regulators. MAN1 interacts with R-Smads, germ-cell-less (GCL), Barrier to autointegration factor (BAF) and BCL-2 associated transcription factor (BTF). Emerin interacts with BAF, BTF, GCL, HDACs, the nuclear corepressor (NCoR) complex and with β-catenin. Lap2β binds lamin B, chromatin, GCL and HDACs (reviewed in Wilson & Foisner, 2010).

Genome-NL interactions were mapped using the DamID technology, in which genomic regions that are in molecular contact with the NL become adenine methylated (a modification unknown to higher eukaryotes) due to expression of a fusion protein consisting of *E. coli* DNA adenine methyltransferase (Dam) and Lamin B1 (Guelen *et al*, 2008). Mammalian genomes contain about 1100–1400 lamina-associated domains (LADs), which are regions of ~0.1–10 Mb that specifically associate with the NL. LADs are relatively gene-poor, have a repressive chromatin signature, are often demarcated by specific sequence elements, such as binding sites for the insulator protein CTCF, and are transcriptionally inactive (Fig. 5.2).

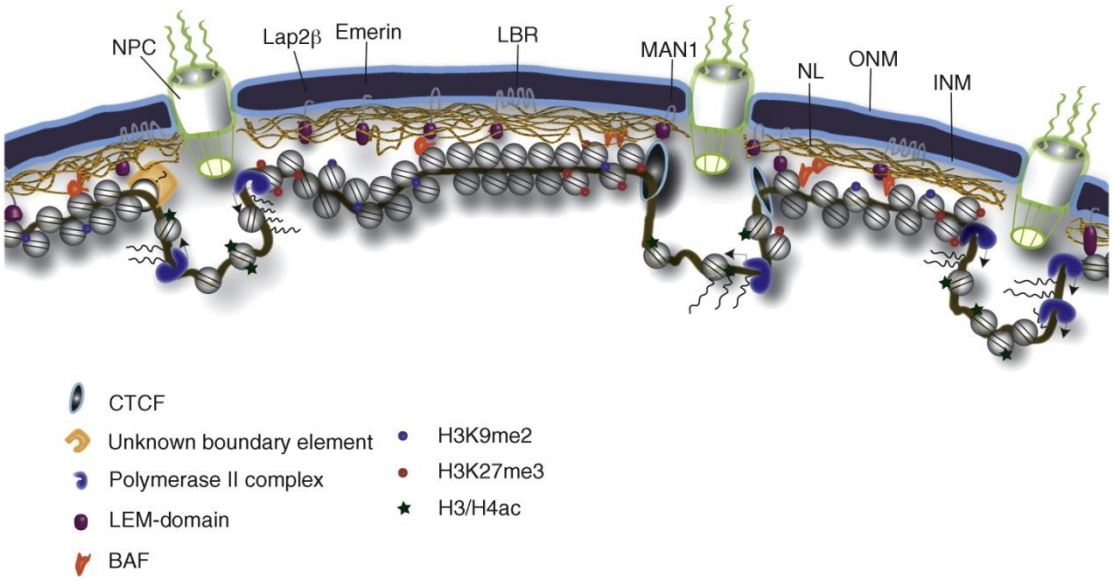

Figure 5.2 Illustration of genome-NL interactions in mammalian cells. Chromatin associated with the NL is marked by specific histone modification, such as H3K9me2 and H3K27me3, and is mostly composed of inactive genes. LADs are bordered by CTCF and possibly by other proteins. NL and nuclear pore complexes (NPCs) form different microenvironments. ONM, outer nuclear membrane; LBR, Lamin B receptor (adapted from Kind & van Steensel, 2010).

The mechanisms that promote LAD-NL interactions are still poorly understood. Recent studies suggest that specific DNA sequences can be recognised by NL-associated DNA-binding proteins. Analysis of two mouse LADs showed enrichment in $(GA)_n$ repeats. The transcriptional repressor cKrox could bind this repeated motif and tether them to the NL by forming a complex with HDAC3 and Lap2β (Zullo *et al*, 2012). However, $(GA)_n$ repeats may not be globally enriched in LADs, indicating that other NL-targeting sequences may exist (Guelen *et al*, 2008). In mammals, constitutive LADs, which are present in all cell types, are characterised by long stretches of DNA of high A/T content (Meuleman *et al*, 2013).

Not only DNA sequence, but also local chromatin properties can provide recognition sites to NL interactions. In human cells, LAD-NL interactions are linked to transcriptional repression and H3K9me2, and are regulated by the H3K9 methyltransferase G9a (Kind *et al*, 2013). Both H3K9me2 and H3K9me3 were found to be involved in the NL-targeting of the human beta-globin locus (Bian *et al*, 2013). Furthermore, treatment with HDAC inhibitors can dissociate mammalian genes from the NP and upregulate their expression (Finlan *et al*, 2008).

It seems that whether the association of a locus with the NL favours, represses or does not affect transcription will depend on the chromatin environment of the locus, how it is recruited to the NL and how it is positioned relative to various nuclear neighbourhoods (Takizawa *et al*, 2008b).

5.1.3 Aims of this Chapter

Reporter transgenes are a powerful tool to reveal mechanisms of gene expression, however the local environment of the integration site may influence their expression (Akhtar *et al*, 2013), interfere with their nuclear positioning and their ability to sit outside of their CTs and contact with sequences from other chromosomes (Brown *et al*, 2006, 2008).

The *FXN-MS2-Luc* cell lines (Chapter 4) are an invaluable tool to study the interplay between *FXN* localisation and expression at the single-cell level. However, *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* transgenes are integrated in chromosome 1. In this Chapter, I analyse the nuclear localisation of GAA expanded *FXN* alleles in their native chromatin environment in chromosome 9, and how their position relative to nuclear compartments is able to impact on *FXN* repression. Furthermore, I analyse whether GAA expansion-mediated relocation of the *FXN* locus to the NL plays a role in FRDA. Here, I investigate the link between expanded *FXN* positioning at the NL and GAA-mediated transcriptional repression in FRDA patient-derived cells. Finally, I elucidate the effect of the mutation at the initiation and elongation stages of *FXN* transcription at single-cell resolution.

5.2 Results

5.2.1 Expanded *FXN* alleles localise preferentially close to the NP in carrier cells

I first analysed *FXN* localisation in its native genomic environment in carrier-derived lymphoblastoid cells (GM14519) carrying 9 and 1285 GAA repeats in each *FXN* allele, respectively. Each carrier cell exhibited a 55% reduction in mRNA compared to healthy cells (Fig. 5.3). Co-detection of a DNA probe for *FXN* with a $(GAA)_{15}$ or $(TTC)_{15}$ oligo signal allowed discrimination of the expanded allele and a direct intracellular comparison of the normal and the expanded *FXN* allele, with distance measurements for each allele to the NP by 2D (Fig. 5.4A) and 3D (Fig. 5.4B) FISH (Meaburn & Misteli, 2008; Takizawa *et al*, 2008a). I found the expanded *FXN* allele localised preferentially closer to the NP than the normal allele.

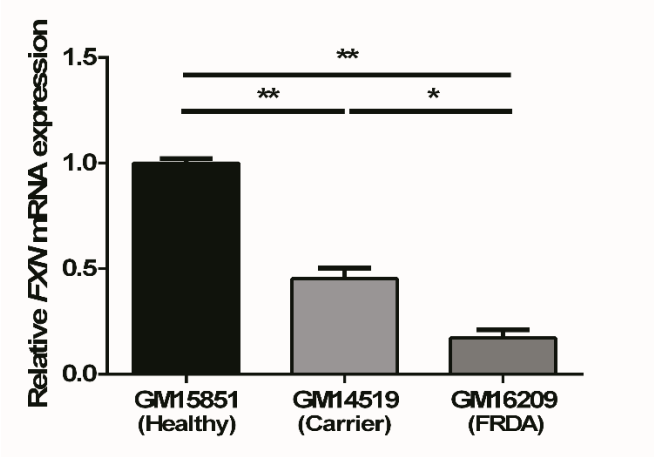

Figure 5.3 *FXN* mRNA expression in healthy, carrier and FRDA patient-derived cells. *FXN* mRNA expression was determined by qRT-PCR using *FXN* primers. Data were normalised to *β-actin*. **P* < 0.05, ***P* < 0.01, as determined by one-way ANOVA with Tukey's multiple comparisons test. Data are mean ± SEM (*n* = 3).

5.2.2 *FXN* gene positioning relative to nuclear subdomains in carrier cells

To further analyse the observed difference in nuclear positioning, I then assessed *FXN* association with specific nuclear subdomains known to either (i) suggest the transcriptional status of a gene, such as nuclear SC35-enriched speckles and CTs, or (ii) interact with heterochromatin, such as the nucleolus and the NL (Shopland *et al*, 2003; Chambeyron & Bickmore, 2004; Brown *et al*, 2006; Zhang *et al*, 2007; Reddy *et al*, 2008; Zhao *et al*, 2009). *FXN* never looped out of chromosome 9 nuclear territory in 2D FISH experiments (Fig. 5.5A), demonstrating that this does not account for the observed difference in nuclear localisation. I found no evidence suggesting the expanded and normal *FXN* alleles associated differently with nuclear speckles (Fig. 5.5B) or with heterochromatin surrounding nucleoli (Fig. 5.5C). However, when I analysed *FXN* positioning at the NL, I found an increased number of expanded *FXN* alleles contacting with the NL when compared with the normal allele (Fig. 5.5D). Taken together, these data suggest GAA repeat expansions increase *FXN* positioning at the NL in carrier-derived cells.

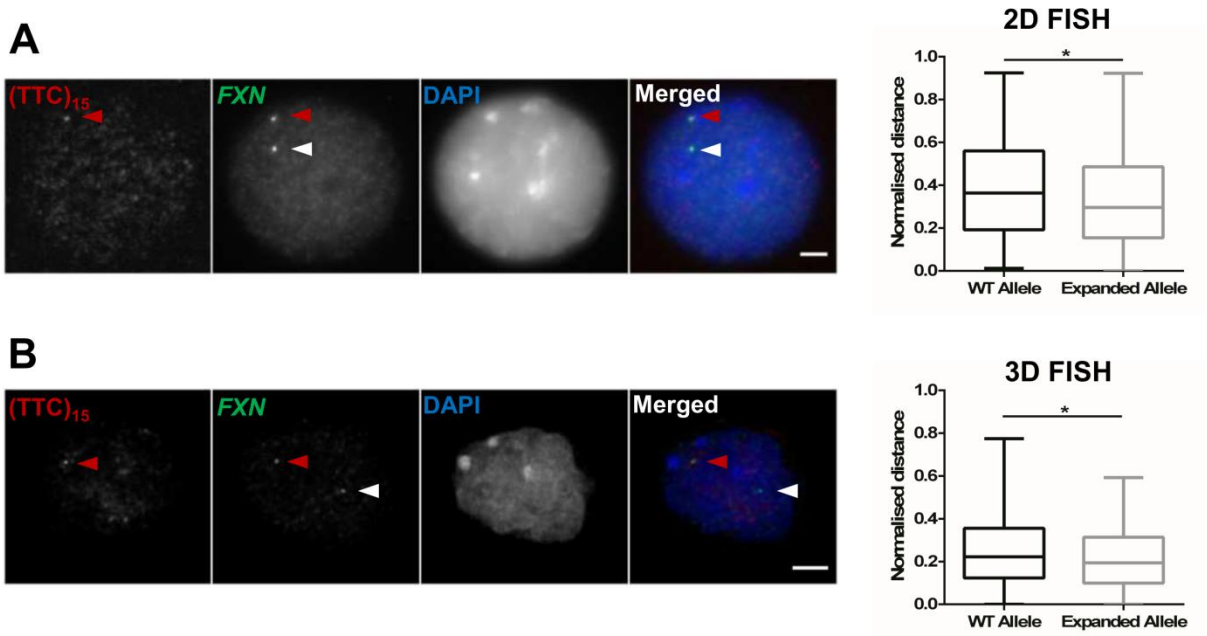

Figure 5.4 Expanded *FXN* alleles localise closer to the NP in carrier cells. Each allele was detected with a DNA probe for *FXN* (green in merged image) and the expanded allele was identified by co-localisation with a $(GAA)_{15}$ or $(TTC)_{15}$ oligo signal (red) (left). Nuclei were identified by DAPI staining (blue). Arrows indicate normal (white) and expanded (red) alleles. (**A**) 2D FISH analysis showed that the expanded *FXN* allele (1285+50 GAA repeats) localised preferentially closer to the NP than the wild-type (WT) (9 GAA repeats) (**P* = 0.0153, as determined by Wilcoxon matched-paired two-tailed test; data are mean ± SEM from three independent hybridisations, *n* = 147). (**B**) 3D FISH analysis confirmed the results obtained by 2D FISH (**P* = 0.0254, as determined by Wilcoxon matched-paired two-tailed test; data are mean ± SEM from three independent hybridisations, *n* = 107). (**A-B**) Distance measurements between FISH signals and the NP were measured and normalised by the local ray between the nuclear centre and periphery passing through the FISH signal (right). Images are maximum projections of Z stacks covering the signals. Scale bar 2 µm.

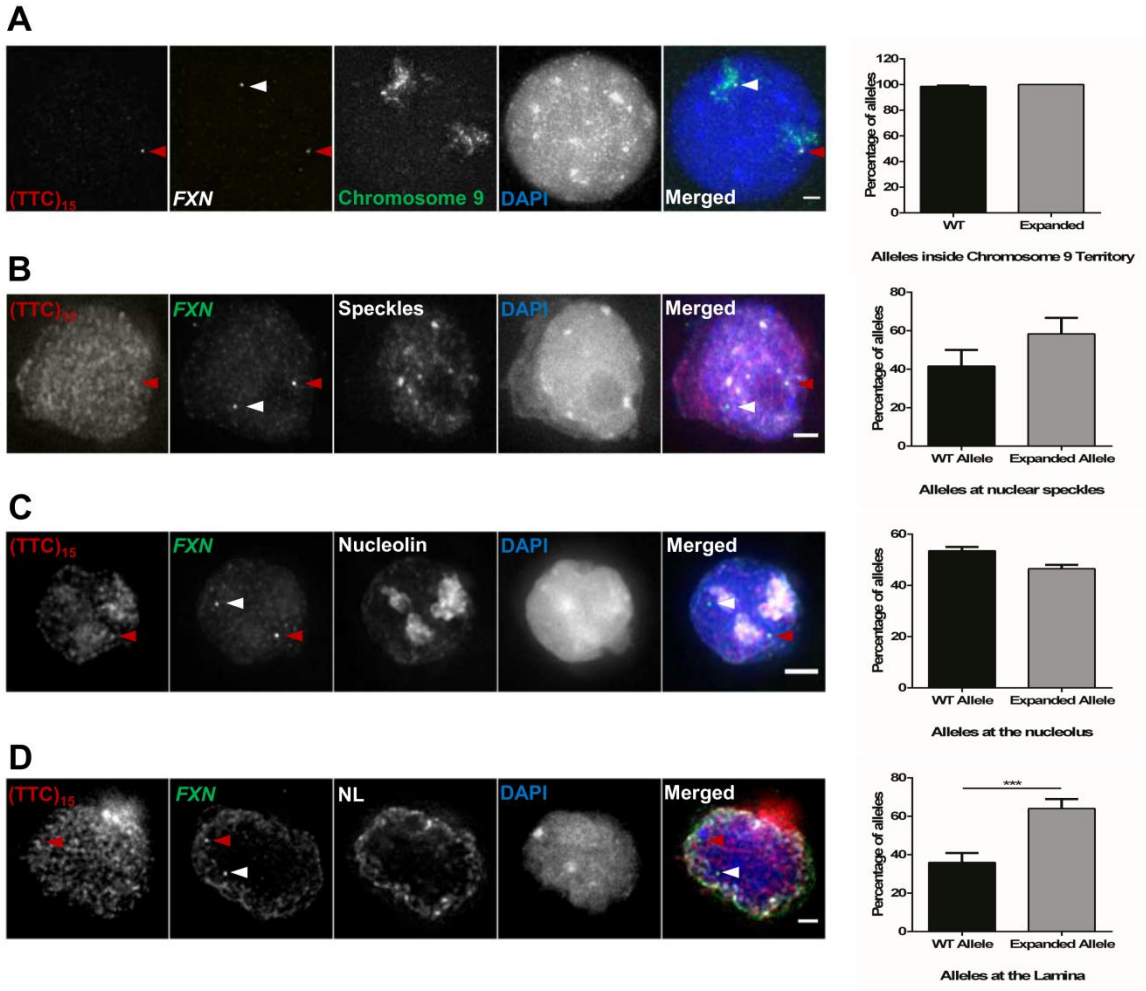

Figure 5.5 Expanded *FXN* allele preferentially sits at the NL in carrier cells. Each allele was detected with a DNA probe for *FXN* (green in merged image) and the expanded allele was identified by co-localisation with a $(GAA)_{15}$ or $(TTC)_{15}$ oligo signal (red) (left). Nuclei were identified by DAPI staining (blue). Arrows indicate normal (white) and expanded (red) alleles. (**A**) *FXN* (white) did not loop out of chromosome 9 territories (green) in 2D FISH experiments. Data are mean ± SEM, *n* =136 cells from three independent hybridisations. (**B**) 3D Immuno-FISH detection of *FXN* (green) association with nuclear SC35-enriched speckles showed no significant difference between alleles, as determined by Fisher's two-tailed exact test ($P = 0.4942$, $n = 86$ cells from two independent hybridisations). (**C**) 3D Immuno-FISH detection of *FXN* association with the nucleoli showed no significant difference between alleles, as determined by Fisher's two-tailed exact test ($P = 0.3607$, $n = 156$ cells from two independent hybridisations) (**D**) 3D Immuno-FISH detection of *FXN* association with the NL (pseudocoloured white). Comparison of the proportion of alleles contacting with the NL revealed the mutation induced a significant increase in co-localisation with the NL, as determined by Fisher's two-tailed exact test (***$P = 0.0005$). Data are mean ± SEM from two independent hybridisations (n = 166). Images are maximum projections of z-stacks covering the signals. Association of FISH signals with subnuclear domains or chromosome territories was analysed using ImageJ: image stacks were analysed manually and association was determined by eye when *FXN* FISH signals were touching or superimposing the signal from a specific subnuclear domain at a given z-section. Scale bar 2 µm.

5.2.3 NL is a key player in *FXN* transcriptional impairment and silencing

I next analysed *FXN* localisation at the NL in healthy (GM15851) and patient-derived cells (GM16209) using Immuno-FISH (Fig. 5.6A). The patient cell line contains 2 *FXN* alleles carrying 800 GAA repeats each and was chosen to avoid differing behaviour of each locus due to different repeat size. Eight-hundred GAA repeats decreased mRNA levels by ~83% (Fig. 5.3). The distribution of cells with *FXN* FISH signals contacting with the NL was significantly different in the 2 cell lines (Fig. 5.6B). I found one or both *FXN* alleles localised at the NL in a small proportion of healthy cells (~37% and ~14%, respectively), with ~49% of cells containing both *FXN* alleles in the nucleoplasm. In contrast, the majority of FRDA cells presented both or one allele contacting with the NL (~43% and ~36%, respectively), with only 21% of FRDA cells containing both *FXN* alleles in the nucleoplasm. When data were analysed at allele level instead of cell level, ~60% of expanded *FXN* alleles were found at the NL compared to only ~33% of normal *FXN* alleles (Fig.5.6B). These data suggest the GAA expansion increases the probability of an allele to be found at the NL in FRDA cells, in agreement with results obtained with *FXN-GAA-MS2-Luc* (Fig. 4.17A) and carrier cells (Fig. 5.5D).

To further dissect *FXN* repression, I analysed nascent *FXN* RNA synthesis at a single-cell level, using probes hybridising with the upstream region of the GAA repeats (Fig. 5.6C). Figure 5.6D shows ~46% of healthy cells contained 2 *FXN* alleles expressing, and ~30% and ~24% of cells with either one or no expressing *FXN* alleles, respectively. I found a higher proportion of FRDA cells (51%) containing both *FXN* alleles not expressing, followed by ~27% and ~22% of cells with either one or both *FXN* alleles expressing, respectively. Analysing these data at allele level instead of cell level, I found ~61% of normal *FXN* alleles were active compared to only ~35% of expanded *FXN* alleles (Fig. 5.6D). These data suggest the mutation increases the probability of an allele being silenced. Collectively, data from Figures 4.17A, 4.18C, 4.20D, 5.5D, 5.6B and 5.6D indicate a clear link between expanded *FXN* positioning at the NL and GAA-mediated transcriptional repression. Furthermore, these results confirm the existence of biological variability within cell populations and highlight the importance of single-cell studies (Levsky & Singer, 2003).

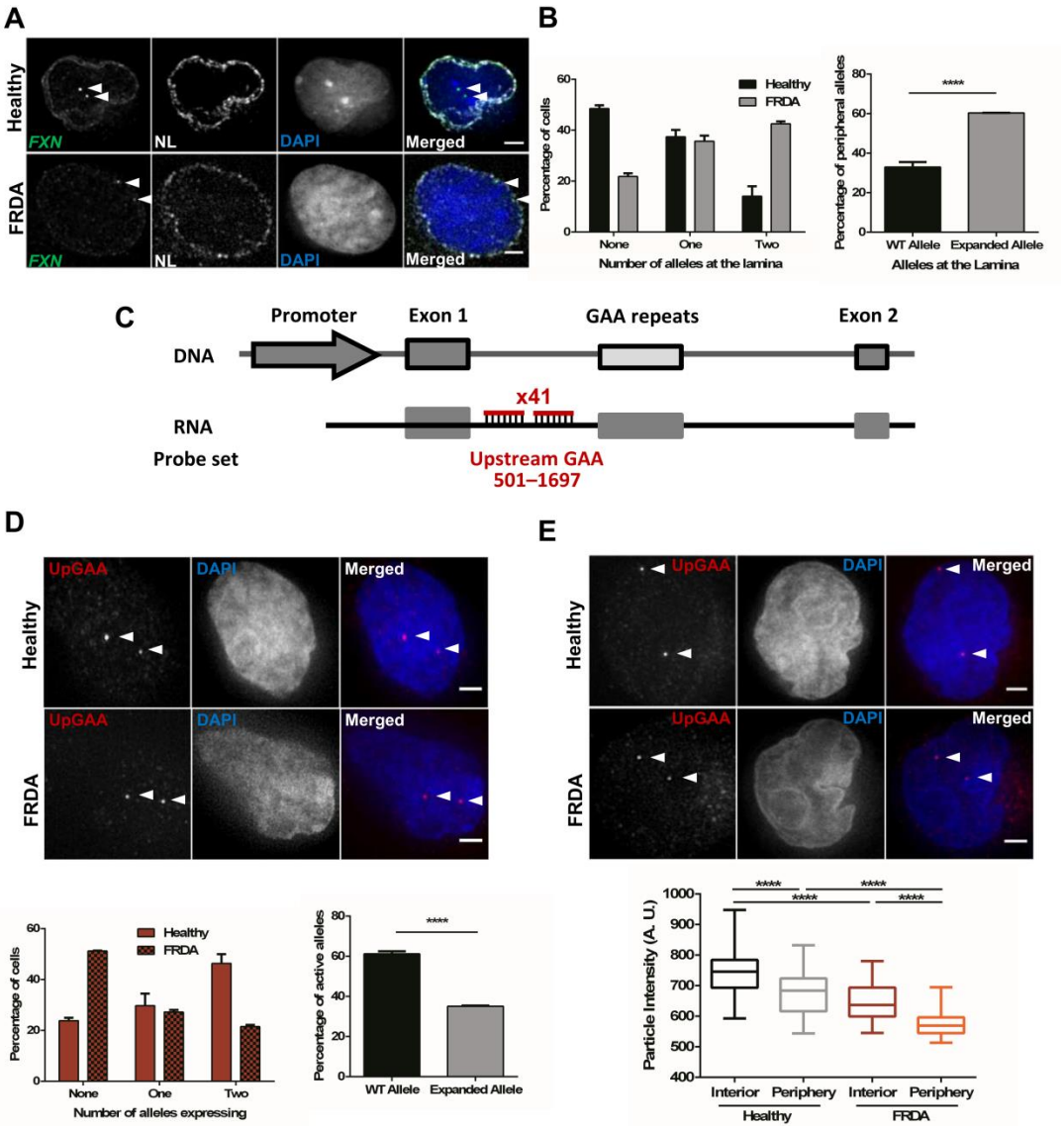

Figure 5.6 GAA repeat expansion increases *FXN* localisation at the NL, reduces the number of active *FXN* alleles and downregulates transcription from active *FXN* in FRDA cells. (**A**) 3D Immuno-FISH detection of *FXN* loci association with the NL in healthy (top images) and patient-derived (bottom images) cells. Each allele was detected with a DNA probe for *FXN* (green in merged image), and the NL was detected with an antibody for Lamin B1 (white). Nuclei were identified by DAPI staining (blue). Arrows indicate alleles. Association of FISH signals with the NL was analysed using ImageJ: image stacks were analysed manually and the association was determined by eye when FXN FISH signals were touching or superimposing the signal from the NL at a given z-section. (**B**) Histogram (left) showing the distribution of cells with *FXN* FISH signals contacting with the NL was significantly different in the two cell lines, as determined by Chi-squared test ($P < 0.0001$, $\chi^2 = 63.056$, d.f. = 2). When data were analysed at allele level instead of cell level (right), the proportion of *FXN* alleles positioned at the NL is significantly different when comparing the localisation of the expanded and normal *FXN* alleles (****$P < 0.0001$, as determined by Fisher's two-tailed exact test). (**C**) Schematic representation showing the relative position to TSS and labelling of the DNA oligo probe set used in nascent RNA FISH experiments. (**D**) Nascent RNA FISH detection of *FXN* expression in healthy and patient-derived cells (top). Each allele was detected with a mix of 41 DNA oligos hybridizing with the region upstream of the GAA repeats (pseudocoloured red in merged image). Histogram (bottom left) showing the distribution of cells with RNA *FXN* FISH signals was significantly different in the two cell lines, as determined by Chi-squared test ($P < 0.0001$, $\chi^2 = 75.424$, d.f. = 2). When data were analysed at allele level instead of cell level (bottom right), the proportion of active *FXN* alleles is significantly different when comparing the expanded and normal *FXN* alleles (****$P < 0.0001$, as determined by Fisher's two-tailed exact test). Data are mean ± SEM from three independent hybridizations (n = 187–202 cells). (**E**) Nascent RNA FISH detection of *FXN* expression in healthy and patient-derived cells (top) showing different signal intensities of internal versus peripheral nascent RNA loci. Detection was performed as described in (D). Box plot (bottom) showing the RNA FISH spot intensity of active *FXN* alleles above the threshold was significantly lower in the interior and at the periphery in FRDA versus healthy nuclei (****$P < 0.0001$) and significantly lower when alleles were expressing at the periphery than in the nuclear interior in the two cell lines, as determined by one-way ANOVA with Tukey's multiple comparisons test (****$P < 0.0001$, n = 245 signals). Images are maximum projections of z-stacks covering the signals. Scale bar: 2 µm.

To further elucidate the interplay between the repressive environment at the NL and *FXN* repression, I analysed RNA FISH spot intensities from active *FXN* alleles in Figure 5.6D as an indication of the quantity of transcript being synthesized at each transcription site relative to its interior or peripheral nuclear position (Fig. 5.6E). The signal variability of the technique was assessed by comparing RNA FISH intensities between replicates. Spot intensities were comparable among replicates (Fig. 5.7A) and background intensity was not significantly different in nuclei of healthy and FRDA patient cells (Fig. 5.7B). I found that *FXN* RNA FISH spot intensity was significantly lower when *FXN* alleles were expressed in the NP compared to the interior in both healthy and FRDA cells. This indicates the NP was not incompatible with *FXN* expression, since *FXN* RNA FISH signals could be seen at the periphery in both healthy and FRDA cells, albeit at lower intensities. Strikingly, RNA FISH spot intensity in FRDA cells was significantly lower than in healthy cells both in the interior and at the periphery. When collectively considering the data in Figures 5.5D, 5.6B and 5.6D, several factors seem to combine to diminish transcription from the expanded *FXN* gene. The expanded *FXN* gene has a significantly increased likelihood of being positioned at the periphery (Figs. 5.5D and 5.6B) and markedly reduced probability of being transcribed at all (Fig. 5.6D). Crucially, when transcription does occur at the expanded GAA-*FXN* locus, it is at a significantly reduced level both in the nucleoplasm and especially at the periphery (Fig. 5.6E). The NP can therefore be seen as a generally less supportive environment for *FXN* transcription, but the combined effect of GAA expansion with peripheral relocation results in a catastrophic reduction in transcriptional output.

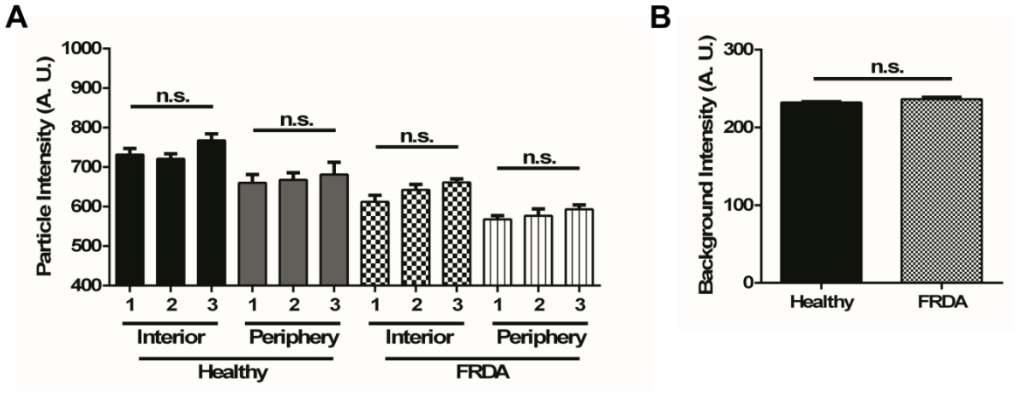

Figure 5.7 Nascent FXN RNA FISH spot intensities in healthy and FRDA patient cells. (**A**) *FXN* RNA FISH spot intensities of active *FXN* alleles in Figure 4E were compared among replicates and shown no significant differences, as determined by one-way ANOVA. Data are mean ± SEM from three independent hybridisations (*n* = 245 signals). (**B**) *FXN* RNA FISH background intensity was not significantly different in nuclei of healthy and FRDA patient cells, as determined by unpaired two-tailed Student's t-test. Data are mean ± SEM from three independent hybridisations (*n* = 245 signals).

5.2.4 Expanded GAA repeats impair preferentially *FXN* transcription initiation in single FRDA cells

To analyse the effect of the GAA expansion at the initiation and elongation stages of *FXN* transcription at single-cell resolution, probes hybridising with the promoter after the transcription start site (TSS) and regions upstream and downstream of GAA repeats were included in the RNA FISH analysis (Figs. 5.8A and B). Healthy *FXN* alleles showed the expected decrease in transcription elongation throughout the analysed regions (Fig. 5.8C and D), confirming previous qRT-PCR analysis (Kim *et al*, 2011; Groh *et al*, 2014). However, when compared to healthy alleles, expanded *FXN* alleles showed a ~19% decrease in RNA levels at the TSS site, demonstrating a marked reduction in *FXN* transcription initiation (Fig. 5.8C). A ~20% decrease at the region upstream of GAA repeats and a ~22% decrease at the downstream region of GAA repeats were also observed. I then normalised the data in Figure 5.8C to exclude the effect of (i) transcription initiation at the TSS and (ii) elongation through the regions upstream of GAA repeats, and found that expanded *FXN* alleles showed a small but significant decrease in *FXN* mRNA synthesis only downstream and not upstream of the GAA repeat expansion when compared with the same regions in healthy alleles (Fig. 5.8D), demonstrating that transcription elongation through expanded GAA repeats is also affected. Taken together, these data indicate GAA repeat expansions diminish *FXN* mRNA synthesis predominantly by impairing *FXN* transcription initiation, but also by impairing elongation through the expansion.

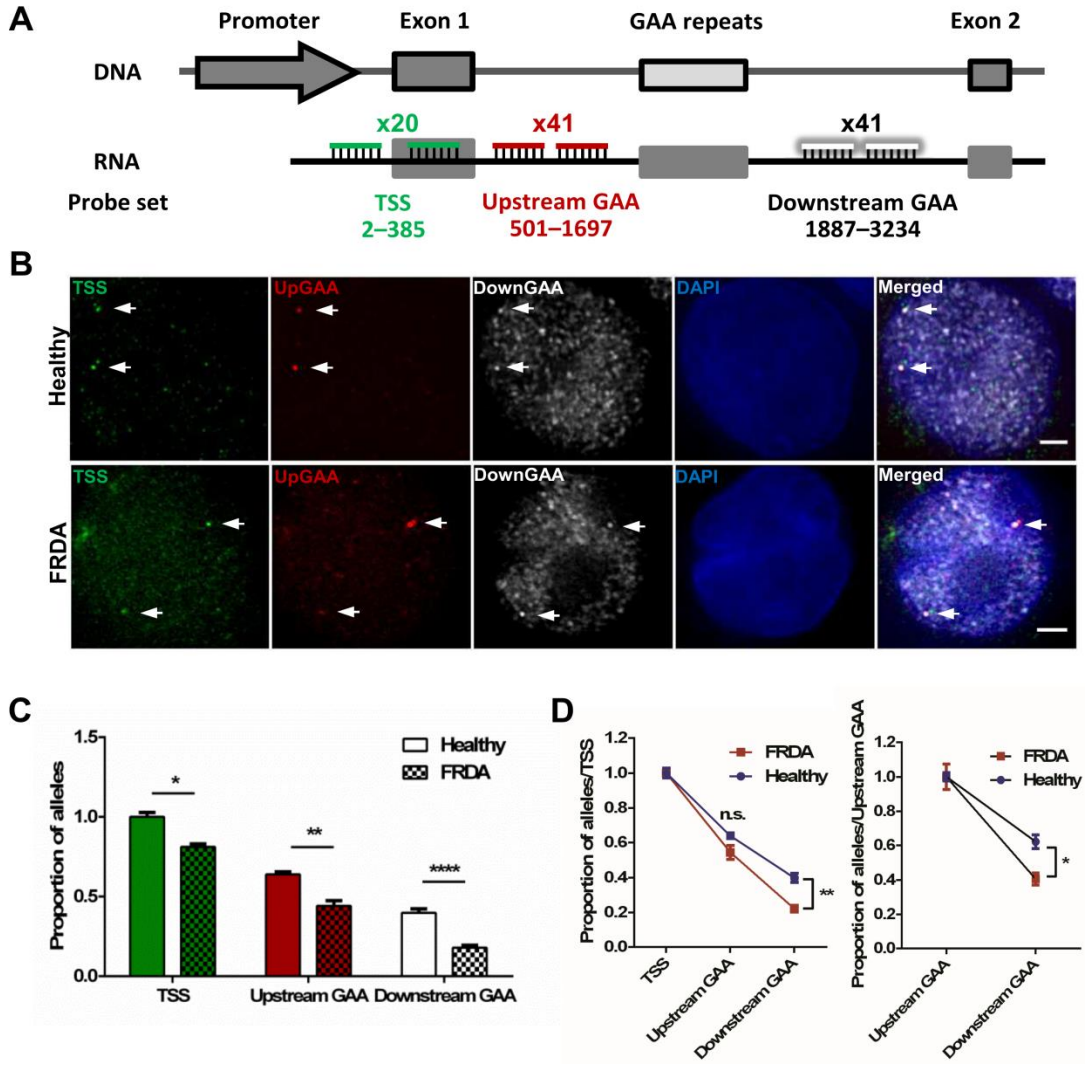

Figure 5.8 GAA repeat expansion disrupts transcription initiation and elongation in FRDA cells.

(**A**) Schematic representation showing the relative position to TSS and labelling of the 3 DNA oligo probe sets used to detect transcription initiation and elongation in RNA FISH experiments. (**B**) *FXN* RNA FISH detection of transcription initiation after the TSS (probe set in green) and elongation through the regions upstream (pseudocoloured red) and downstream (pseudocoloured white) of the GAA repeats. Nuclei were identified by DAPI staining (blue). Arrows indicate alleles. Images are maximum projections of Z stacks covering the signals. Scale bar 2 µm. (**C**) Data were normalised to the TSS of healthy cells. When compared to the same region in healthy alleles, expanded *FXN* alleles showed a ~19% decrease at the TSS, a ~20% and a ~22% decrease at the regions upstream and downstream of GAA repeats, respectively. This data demonstrates GAA repeat expansions reduce *FXN* transcription initiation by 19% and elongation downstream of expanded GAA repeats by 3% (**P* = 0.0119, ***P* = 0.003, *****P* < 0.0001, as determined by Fisher's two-tailed exact test, *n* = 218-226 cells per line). Data are mean ± SEM from three independent FISH experiments. (**D**) Data of each line in (C) were normalised to their TSS, thus excluding the effect of transcription initiation (left) and elongation through the region upstream of GAA repeats (right). Expanded *FXN* alleles show a significant difference in the region downstream of the GAA repeats when comparing with the same region in healthy alleles (**P* < 0.05, ***P* < 0.01, as determined by two-way ANOVA with Sidak's multiple comparisons test). Data are mean ± SEM from three independent FISH experiments.

5.3 Discussion

Nuclear organisation and chromatin structure are important for the regulation of gene expression and, if altered, may cause disease. The genome is organised in a non-random manner and gene expression is regulated in specific and spatially defined sites, where transcription activation or repression are favoured (Misteli, 2007). In FRDA, GAA expansion-mediated transcriptional dysregulation occurs due to the generation of unusual DNA structures and epigenetic changes that induce a non-permissive chromatin configuration for *FXN* expression. In this Chapter, I used a carrier-derived cell line as a model to determine where GAA expanded *FXN* alleles are found in the nucleus at single-cell resolution, and how their location may contribute to *FXN* repression.

The observed 55% reduction in *FXN* mRNA in carrier cells reflects the transcriptional impairment of the expanded *FXN* when compared with the healthy allele. By comparing the radial position of the expanded and normal *FXN*, I showed that the expanded *FXN* allele occupies a different intranuclear position from the normal allele. A possible explanation could be that, given the expanded *FXN* heterochromatic state is coupled with gene repression, the normal *FXN* allele can unfold out of its chromosome 9 territory while being transcribed. However, neither *FXN* alleles sit outside of chromosome 9 territory, irrespective of their different local chromatin condensation and transcriptional status. Active genes can sit inside their CTs and transcribe from within domains of closed chromatin (Gilbert *et al*, 2004; Scheuermann *et al*, 2004). Looping out of the territory of a chromosome is dependent of several features of the chromatin surrounding a gene, such as gene density, chromatin fibre structure, transcriptional activity across the region, and location on the chromosome (Brown *et al*, 2006; Gilbert *et al*, 2004). The *FXN* locus is located on chromosome 9q21.11, an A/T-rich and gene-poor G-dark band close to the centromere (Fig. 5.9). Sitting within the CT may be a reflection of the surrounding chromatin context of the *FXN* locus rather than its transcriptional activity. Furthermore, a possible disruption of the localisation of chromosome 9 territory due to the presence of the expanded *FXN* allele was not readily perceptive given (i) the two chromosomes were indistinguishable, and (ii) the subtle movement of the expanded *FXN* locus when compared to the normal *FXN* allele. Moreover, the heterochromatinised GAA expansion is minute when compared with the whole CT and unlikely to affect the total positioning of a chromosome normally located peripherally.

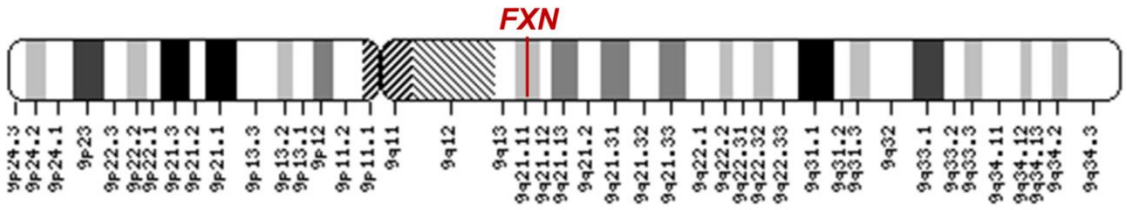

Figure 5.9 Chromosome location of the *FXN* locus. Ideogram of the G-banding pattern of chromosome 9 at the 850 band resolution. The *FXN* locus is located at 9q21.11, an A/T-rich and gene-poor G-dark band (taken from the NCBI Human Genome resource site).

Active genes can be found associated with nuclear speckles (Shopland *et al*, 2003; Brown *et al*, 2006), which are distributed predominantly throughout the nucleoplasm and less frequently at the NP. Moreover, gene-rich chromosome bands contact more often with speckles than gene-poor G-dark bands. The fact that the *FXN* gene is on a gene-poor G-dark band of chromosome 9, with its territory located towards the NP in lymphoblast cells (Boyle *et al*, 2001) may explain the reduced association of the *FXN* locus with nuclear speckles.

The NL and the nucleolus are surrounded by repressive chromatin and genes located in their proximity are often transcriptionally inactive. Distance measurements between the *FXN* locus and the nuclear centre or periphery place the expanded *FXN* locus towards the NP and the normal allele towards the nuclear interior. *FXN* alleles did not contact with the heterochromatin surrounding nucleoli differently. In contrast, I found the expanded *FXN* allele sitting at the NL more often than the normal allele in carrier cells. This is supported by my previous findings indicating that expanded GAA repeats increase the contact of the transgene with the NL in *FXN-GAA-MS2-Luc* cells (Chapter 4) and demonstrates that *FXN* repositioning occurs irrespective of its chromatin environment.

Gene repositioning is dependent on breakdown and reformation of the nuclear envelope during mitosis, with chromatin domain position being established early in G1 (Thomson *et al*, 2004). The variation found in expanded *FXN* repositioning to the NL in carrier cells may reflect the use of unsynchronised carrier cells in this study.

I next analysed the link between *FXN* localisation at the NL and *FXN* silencing in FRDA. I observed the same abnormal repositioning to the NL in FRDA patient cells and showed that this tightly correlates with a marked decrease in the number of actively expressing *FXN* alleles. Furthermore, those few active expanded *FXN* alleles located at the NL express at a significantly lower level than the alleles located in the interior of the nucleus.

Due to limited resolution of light microscopy, FISH experiments shown here cannot distinguish whether the expanded *FXN* allele establishes a functional molecular interaction with the NL that leads to *FXN* repression, or if it is merely in its proximity. However, lamina-associating sequences are enriched in an extended GA dinucleotide sequence which acts as an INM and NL-anchoring motif and is associated with HDACs and transcriptional repressors (Zullo *et al*, 2012). Studies suggest that GAA repeat expansions induce an enrichment of HP1 (De Biase *et al*, 2009), G9a methyltransferase (Groh *et al*, 2014) and HDACs class I (Herman *et al*, 2006) and III (Chan *et al*, 2013) on the *FXN* locus. Consistently, histones in LADs are hypoacetylated, indicating a high local HDAC activity. Furthermore, several NL components interact with HDACs, including Lap2β and Emerin (Kind & van Steensel, 2010). The results from Chapter 4, showing a significant movement from the NL of only the expanded *FXN* alleles after treatment with HDAC inhibitors, supports that the NL represses *FXN* expression partly by deacetylation of histones. The NP is also enriched in HP1, which may contribute to NP targeting due to its interaction with the LaminB receptor (Ye *et al*, 1997).

A second histone modification that appears to be involved in gene regulation at the NL is H3K9me2. Peripheral siting of chromatin after mitosis is largely determined by a stochastic mechanism and directly linked to cellular or allelic variation in H3K9me2 levels where G9a methyltransferase acts as a regulator of NL contacts (Kind *et al*, 2013).

Guenther and colleagues showed that most genes considered inactive, because they produce no detectable transcript, still experience transcription initiation, indicating a post-initiation regulation (Guenther *et al*, 2007). Detection of *FXN* nascent transcript at single-cell resolution showed that both normal and expanded *FXN* alleles have a decrease in transcription elongation throughout the locus, indicating that occasionally the RNAPII may initiate transcription with no production of a full-length transcript. However, here I showed that expanded GAA repeats impair expression predominantly at the initiation stage of transcription. Interestingly, there are genes, such as clusters of silent olfactory receptor genes, that do not experience transcription initiation and do not contain nucleosomes enriched in H3K4me3 (Guenther *et al*, 2007). These clusters form heterochromatic foci where expression is repressed and frequently associate with the NL (Peric-Hupkes *et al*, 2010).

CHAPTER 6

General discussion and future directions

In FRDA, expanded GAA repeats in intron 1 of the *FXN* gene reduce steady-state *FXN* mRNA levels through a poorly understood mechanism. GAA expansion-mediated transcriptional dysregulation is thought to occur due to the generation of unusual DNA structures surrounding the GAA repeats, such as triplexes or sticky DNA (Sakamoto *et al*, 1999; Bidichandani *et al*, 1998), R-loops (Grabczyk *et al*, 2007; Groh *et al*, 2014) and heterochromatin (Saveliev *et al*, 2003; Herman *et al*, 2006), which interfere with RNAPII elongation (Punga & Bühler, 2010; Kim *et al*, 2011). Although not universally accepted, some reports show that these epigenetic changes also spread upstream towards the *FXN* promoter, inducing a non-permissive chromatin configuration for transcription initiation, altering nucleosome positioning and preventing insulator protein CTCF binding (De Biase *et al*, 2009; Kumari *et al*, 2011; Chutake *et al*, 2014; Groh *et al*, 2014). In order to better understand the underlying pathogenesis and ultimately to design effective therapies for FRDA, it is important to generate cellular models that recapitulate the repressive hallmarks of the disease while providing efficient ways to quantify *FXN* expression.

The *FXN-GAA-Luc, FXN-Luc, FXN-GAA-MS2-Luc* and *FXN-MS2-Luc* reporter cell lines, described in this Thesis, have been specifically designed to allow a direct single-allele comparison of the effect of normal or expanded GAA repeats on *FXN* expression within the genomic context of the *FXN* locus. This was achieved by: i) using the whole *FXN* locus with its native promoter, introns, exons and all elements required for physiological transgene expression; and ii) integrating a single copy of each BAC vector into a specific FRT site in chromosome 1 of HEK FRT cells to exclude confounding effects on *FXN* expression due to random site of integration. However, the *FXN-Luc* and *FXN-MS2-Luc* models have been generated with distinct purposes. In the *FXN-GAA-Luc* cell lines, the presence of ~310 GAA repeats recapitulates the GAA repeat-mediated *FXN* repression and the epigenetic hallmarks of FRDA. Therefore, the *FXN-GAA-Luc* model is an excellent tool to study the molecular mechanisms underlying FRDA as well as providing a platform for the screening of *FXN* upregulating compounds. However, studies using the *FXN-GAA-Luc* model, similar to other FRDA reporter systems studies to date, only provide the probable state of the *FXN* gene as observations come from experiments in which the outputs of bulk cell cultures are averaged. Consequently, the *FXN-GAA-MS2-Luc* model was generated to dissect the silencing mechanism of FRDA *in situ*, allowing the visualization and analysis of *FXN* localization and repression in both fixed and living single cells using the MS2 system.

Through these experiments, I demonstrate that the GAA expansion-mediated impairment of *FXN* expression is the result of a complex interplay of multiple factors contributing to different extents to decreased *FXN* transcriptional activity (Fig. 6.1). Among these factors I identify the NL, a location known to be a generally repressive environment, as a novel major player in *FXN* silencing in FRDA. Using a variety of cellular models and a multidisciplinary approach, I found a marked increase in expanded *FXN* positioning at the NL and a concomitant reduction in the number of active *FXN* alleles, suggesting that *FXN* localised at the NL leads to *FXN* silencing. I have also identified a small proportion of expanded *FXN* alleles sitting at the NL which are transcriptionally active, however these show significantly reduced expression when compared with alleles in the nuclear interior.

In my work, I observed two distinct events affecting *FXN* expression. First, I observed that the GAA repeat expansion impairs *FXN* transcription regardless of *FXN* localisation in the nuclear interior or periphery. Second, although the proportion of normal and expanded *FXN* alleles positioned at the NL is remarkably different, I found that the NL reduces the expression of both normal and expanded *FXN* at the NP to a similar extent, demonstrating that the repressive action of the NL is irrespective of a GAA expansion. This leads to a consideration of the order of molecular events occurring at the expanded *FXN* locus. One of the initial steps in *FXN* repression is the binding of nascent *FXN* mRNA to the GAA repeat expansion as the gene is being transcribed, a feature shared with the CGG repeat expansion disease FRAXA (Groh *et al*, 2014; Colak *et al*, 2014). These RNA/DNA hybrids (R-loops) have been previously shown to form on both transgenic and endogenous expanded *FXN* loci in *FXN-GAA-Luc* and FRDA patient-derived lymphoblastoid cell lines, respectively, used in this study (Groh *et al*, 2014). R-loops impair RNAPII elongation and trigger epigenetic changes, such as H3K9me2 repressive marks through recruitment of G9a methyltransferase (Groh *et al*, 2014). The effects of inhibition of HDACs were interesting in this regard. *FXN* transcription from both normal and expanded alleles is upregulated by increased acetylation by similar amounts, yet I observe a significant movement away from the NL of only the expanded alleles. The mechanism of retention at the NL of these expanded repeat structures is therefore highly sensitive to HDAC inhibition. However, the expanded alleles are unable to achieve the same level of upregulation as their normal counterparts. This could be the result of a different extent of histone acetylation achieved by normal and expanded alleles following treatment with HDAC inhibitors, which could be caused by reduced accessibility of

compacted *FXN* alleles. However, the observed difference in *FXN* upregulation could also imply that other structural changes such as R-loop formation contribute to transcriptional dysregulation and that the trigger event is induced by structural and epigenetic changes to the DNA followed by an increased likelihood of retention at the NL. Further experiments to clarify the order of events could involve cell treatment with other compounds, for example DNA sequence-specific polyamides to destabilise triplexes, RNase H1 to resolve R-loops, and G9a inhibitors to interfere with *FXN* retention at the NL. Treatment with polyamides (Burnett *et al*, 2006) and RNAse H1 (Groh *et al*, 2014) have previously resulted in upreagulation of *FXN* expression in FRDA lymphoblastoid cells. Recent studies using the G9a inhibitor BIX-01294 showed a significant reduction in the levels of H3K9me2 chromatin mark, but no effect on *FXN* mRNA levels (Punga & Bühler, 2010; Groh *et al*, 2014).

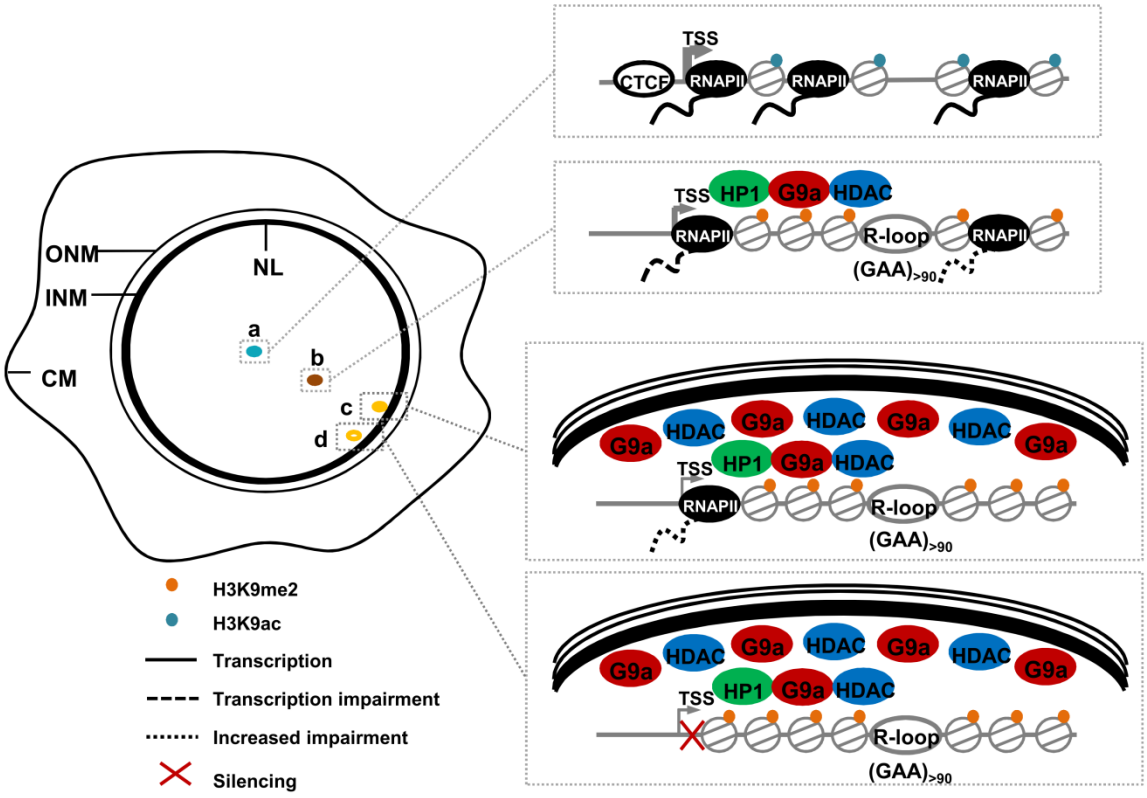

Figure 6.1 Model of GAA repeat-mediated *FXN* repression and the repressive role of NL in *FXN* transcriptional impairment. The euchromatic state of normal *FXN* alleles allows physiological levels of expression (**a**). Expanded GAA repeats trigger R-loop formation and locally recruit HP1 (De Biase *et al*, 2009), G9a methyltransferase (Groh *et al*, 2014) and class I (Herman *et al*, 2006) and III (Chan *et al*, 2013) HDACs, causing epigenetic changes that contribute to transcriptional dysregulation (**b**). These GAA expansion-mediated structural and epigenetic changes may lead to *FXN* relocation to the NL—a location known to be a generally repressive environment and enriched in HDAC and histone methyltransferase activity—where further repression (**c**) and silencing (**d**) occur. ONM, outer nuclear membrane; INM, inner nuclear membrane; CM, cell membrane.

It seems plausible that, once a locus adopts an abnormal DNA structure and becomes modified by specific markers, it is more likely to be retained at the NL after mitosis (Schooley *et al*, 2012). There are several mechanisms that could be involved. Genome contacts with the NL during interphase are generally dynamic, but confined to a narrow zone near the NL. Peripheral siting of chromatin after mitosis is largely determined by a stochastic mechanism and directly linked to cellular or allelic variation in H3K9me2 levels where G9a methyltransferase acts as a regulator of NL contacts (Kind *et al*, 2013). The NP is enriched in heterochromatin and the presence of HP1, whose binding is promoted by methylation at lysine 9, may help to regulate the post-mitotic binding of INM proteins (Schooley *et al*, 2012). There are also spatial considerations. The *FXN* locus is located between two LADs — NL-interacting chromosome regions demarcated by the insulator protein CTCF that present heterochromatic histone modifications and low expression levels—in lung fibroblasts (Guelen *et al*, 2008) and various neural cell types (Peric-Hupkes *et al*, 2010). LAD topology has yet to be assessed in the lymphoblastoid cell lines used in this study, but as different cell types present remarkable LAD similarities (Peric-Hupkes *et al*, 2010), it is tempting to speculate that *FXN* is also in an interLAD in lymphoblastoid cells. Given the severe depletion of CTCF in the 5' UTR of the expanded *FXN* in FRDA (De Biase *et al*, 2009), I speculate that the boundary between *FXN* and the LAD located upstream is inefficient, leading to the inclusion of expanded *FXN* inside the LADs and increased proximity to the NL, possibly through recruitment of NL-associated HDACs and G9a methyltransferase (Groh *et al*, 2014). It would be interesting to explore the LAD layout over this region using the DamID technology (Guelen *et al*, 2008) in FRDA versus healthy individuals, and whether interruption of the boundary CTCF elements can allow consolidation of the two flanking LADs. A further spatial contribution could be that even normal *FXN* can stochastically contact the NL, considering the intranuclear mobility of chromatin (Chubb *et al*, 2002) and the *FXN* native location on the usually peripheral chromosome 9. Indeed I observe that transcription from the normal allele is downregulated to an equivalent degree when located at the NP. *FXN* retention may result in a molecular interaction with the NL, as lamina-associating sequences are enriched in a GAGA motif which is bound by the transcriptional repressor cKrox associated with HDAC3 and INM Lap2β (Zullo *et al*, 2012). It would be interesting to knockdown cKrox in FRDA patient cells and analyse whether expanded *FXN* alleles at the NL relocalise to the nuclear interior. These events in turn could act through a positive feedback mechanism by mediating further impairment or silencing since the NL is enriched in HDAC and histone methyltransferase

activity. Given the shared features of DNA repeat expansions, I hypothesize that the NL may play a role in heterochromatin-mediated transcriptional dysregulation in other repeat-expansion diseases and it would be interesting to interrogate the nuclear location of other repeat expansion loci.

Although I show a marked degree of repression at the NP, one striking finding is that of transcriptionally active *FXN* alleles sitting at the NL, confirming what others have shown that the NP creates a repressive environment, but is not incompatible with expression (Finlan *et al*, 2008; Kumaran & Spector, 2008). These findings support the existence of NL heterogeneous microdomains composed of nuclear pore complexes and different lamin subtypes (Kind & van Steensel, 2010) able to impact differently on gene expression.

It is widely accepted that expanded GAA repeats cause *FXN* mRNA deficiency via RNAPII elongation impairment. I demonstrate that expanded GAA repeats predominantly disrupt *FXN* transcription initiation at single-cell resolution, providing strong confirmation that the transcriptional repressive effect of the GAA repeat expansion can spread towards the *FXN* promoter (Kumari *et al*, 2011; Chutake *et al*, 2014).

It remains undetermined how far the GAA-mediated repression can span. Indeed, non-coding triplet repeat expansions can mediate gene repression of other genes in *cis* (Klesert *et al*, 1997; Thornton *et al*, 1997). There is evidence suggesting that *PIP5K1B* is downregulated in FRDA lymphocytes and fibroblasts (Bayot *et al*, 2013). The *PIP5K1B* locus is located ~26 kb upstream of *FXN* and encodes phosphatidylinositol 4-phosphate 5-kinase β type I, an enzyme functionally linked to actin cytoskeleton dynamics. This gene is highly expressed in the brain, but also in the heart and lungs to a lesser extent (Volpicelli-Daley *et al*, 2010), thus partially overlapping with tissues where *FXN* is expressed (Campuzano *et al*, 1996). Two other genes, *PRKACG* and *TJP2*, are located immediately upstream and downstream of the *FXN* locus, respectively, and show no differences in expression in FRDA patient cells (Bayot *et al*, 2013). Recently, the most extensive genome-wide map of chromosomal interactions revealed that kidney, lung and liver disease-associated polymorphisms in the *PIP5K1B* and *TJP2* loci interact with the *FXN* promoter in a lymphoblastoid cell line (GM12878), suggesting that these elements may potentially influence *FXN* regulation, either positively or negatively (Mifsud *et al*, 2015). Information on the *FXN* locus interaction in *trans* is not yet available. However, it would be interesting to analyse the expanded *FXN* localisation in FRDA-relevant tissues, such as

cardiomyocytes and sensory neurons, and assess expression levels of potential spatial *FXN*-interacting genes.

Recently, a report showed that R-loops formed by the CGG expanded repeat in FRAXA reduces *FMR1* expression via recruitment of the Polycomb group (PcG) complexes to the *FMR1* locus (Kumari & Usdin, 2014). The PcG protein complexes are present in silenced genes and include Polycomb repressive complexes 1 and 2 (PRC1 and PRC2) (Di Croce & Helin, 2013). Interestingly, recent microarray data from neuronal cells (derived from FRDA patient iPSCs) revealed that the gene encoding the Polycomb group ring finger 2 (PCGF2) is downregulated after treatment with HDAC inhibitor 109, a potent inducer of *FXN* gene expression (Soragni *et al*, 2015). PCGF2 is a component of PRC1. PRC2 catalyses the trimethylation of H3K27, which acts as the docking site for PRC1 (Di Croce & Helin, 2013). Due to enrichment of the repressive hallmark H3K27me3 at the expanded *FXN* locus (De Biase *et al*, 2009; Kim *et al*, 2011; Chan *et al*, 2013), future work may explore whether the PcG proteins may also play a role in *FXN* silencing in FRDA.

Here, I have focused on the nuclear location of and transcription from the expanded *FXN* gene at the single-cell level. I find an increased retention of the expanded *FXN* gene at the NL. I show that a higher proportion of expanded *FXN* alleles are inactive when compared with normal alleles and I have demonstrated reduced transcription from the expanded alleles that are active. I provide a single-cell quantification showing that expanded GAA repeats impair expression predominantly at the initiation stage of transcription, but also block RNAPII elongation. It is the combination of these factors which together reduce the levels of frataxin protein to cause disease. This Thesis provides new mechanistic insights into the molecular causes of FRDA which may extend to other genetic diseases mediated by repeat expansions within regions of non-coding DNA.

REFERENCES

Adinolfi S, Iannuzzi C, Prischi F, Pastore C, Iametti S, Martin SR, Bonomi F & Pastore A (2009) Bacterial frataxin CyaY is the gatekeeper of iron-sulfur cluster formation catalyzed by IscS. *Nat. Struct. Mol. Biol.* **16:** 390–396

Akhtar W, De Jong J, Pindyurin A V., Pagie L, Meuleman W, De Ridder J, Berns A, Wessels LF a, Van Lohuizen M & Van Steensel B (2013) Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell* **154:** 914–927

Al-Mahdawi S, Pinto RM, Ismail O, Varshney D, Lymperi S, Sandi C, Trabzuni D & Pook M (2008) The Friedreich ataxia GAA repeat expansion mutation induces comparable epigenetic changes in human and transgenic mouse brain and heart tissues. *Hum. Mol. Genet.* **17:** 735–746

Al-Mahdawi S, Pinto RM, Ruddle P, Carroll C, Webster Z & Pook M (2004) GAA repeat instability in Friedreich ataxia YAC transgenic mice. *Genomics* **84:** 301–310

Al-Mahdawi S, Sandi C, Mouro Pinto R & Pook MA (2013) Friedreich ataxia patient tissues exhibit increased 5-hydroxymethylcytosine modification and decreased CTCF binding at the FXN locus. *PLoS One* **8:** e74956

Aloria K, Schilke B, Andrew A & Craig EA (2004) Iron-induced oligomerization of yeast frataxin homologue Yfh1 is dispensable in vivo. *EMBO Rep.* **5:** 1096–1101

Anjomani Virmouni S, Ezzatizadeh V, Sandi C, Sandi M, Al-Mahdawi S, Chutake Y & Pook MA (2015) A novel GAA-repeat-expansion-based mouse model of Friedreich's ataxia. *Dis. Model. Mech.* **8:** 225–235

Anjomani Virmouni S, Sandi C, Al-Mahdawi S & Pook MA (2014) Cellular, molecular and functional characterisation of YAC transgenic mouse models of Friedreich ataxia. *PLoS One* **9:** e107416

Babcock M, de Silva D, Oaks R, Davis-Kaplan S, Jiralerspong S, Montermini L, Pandolfo M & Kaplan J (1997) Regulation of mitochondrial iron accumulation by Yfh1p, a putative homolog of frataxin. *Science* **276:** 1709–1712

Bannister AJ & Kouzarides T (2011) Regulation of chromatin by histone modifications. *Cell Res.* **21:** 381–395

Baralle M, Pastor T, Bussani E & Pagani F (2008) Influence of Friedreich ataxia GAA noncoding repeat expansions on pre-mRNA processing. *Am. J. Hum. Genet.* **83:** 77–88

Bayot A, Reichman S, Lebon S, Csaba Z, Aubry L, Sterkers G, Husson I, Rak M & Rustin P (2013) Cis-silencing of PIP5K1B evidenced in Friedreich's ataxia patient cells results in cytoskeleton anomalies. *Hum. Mol. Genet.* **22:** 2894–2904

Ben-Ari Y, Brody Y, Kinor N, Mor A, Tsukamoto T, Spector DL, Singer RH & Shav-Tal Y (2010) The life of an mRNA in space and time. *J. Cell Sci.* **123:** 1761–1774

Bernardi R & Pandolfi PP (2007) Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. *Nat. Rev. Mol. Cell Biol.* **8:** 1006–1016

Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH & Long RM (1998) Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2:** 437–445

Bian Q, Khanna N, Alvikas J & Belmont AS (2013) β-Globin cis-elements determine differential nuclear targeting through epigenetic modifications. *J. Cell Biol.* **203:** 767–783

De Biase I, Chutake YK, Rindler PM & Bidichandani SI (2009) Epigenetic silencing in Friedreich ataxia is associated with depletion of CTCF (CCCTC-binding factor) and antisense transcription. *PLoS One* **4:** e7914

De Biase I, Rasmussen A, Endres D, Al-Mahdawi S, Monticelli A, Cocozza S, Pook M & Bidichandani SI (2007a) Progressive GAA expansions in dorsal root ganglia of Friedreich's ataxia patients. *Ann. Neurol.* **61:** 55–60

De Biase I, Rasmussen A, Monticelli A, Al-Mahdawi S, Pook M, Cocozza S & Bidichandani SI (2007b) Somatic instability of the expanded GAA triplet-repeat sequence in Friedreich ataxia progresses throughout life. *Genomics* **90:** 1–5

Bidichandani S, Garcia C, Patel P & Dimachkie M (2000) Very late-onset Friedreich ataxia despite large GAA triplet repeat expansions. *Arch. Neurol.* **57:** 246–251

Bidichandani SI, Ashizawa T & Patel PI (1997) Atypical Friedreich ataxia caused by compound heterozygosity for a novel missense mutation and the GAA triplet-repeat expansion. *Am. J. Hum. Genet.* **60:** 1251–1256

Bidichandani SI, Ashizawa T & Patel PI (1998) The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. *Am. J. Hum. Genet.* **62:** 111–121

Bolte S & Cordelières FP (2006) A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* **224:** 213–232

Bolzer A, Kreth G, Solovei I, Koehler D, Saracoglu K, Fauth C, Müller S, Eils R, Cremer C, Speicher MR & Cremer T (2005) Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol.* **3:** 0826–0842

Bond CS & Fox AH (2009) Paraspeckles: nuclear bodies built on long noncoding RNA. *J. Cell Biol.* **186:** 637–644

Bourn RL, De Biase I, Pinto RM, Sandi C, Al-Mahdawi S, Pook MA & Bidichandani SI (2012) Pms2 suppresses large expansions of the (GAA·TTC)n sequence in neuronal tissues. *PLoS One* **7:** e47085

Boyle S, Gilchrist S, Bridger JM, Mahy NL, Ellis JA & Bickmore WA (2001) The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum. Mol. Genet.* **10:** 211–220

Branco MR & Pombo A (2006) Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol* **4:** e138

Brodsky AS & Silver PA (2002) Identifying proteins that affect mRNA localization in living cells. *Methods* **26:** 151–155

Van den Broek WJAA, Nelen MR, Wansink DG, Coerwinkel MM, te Riele H, Groenen PJTA & Wieringa B (2002) Somatic expansion behaviour of the (CTG)n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch–repair proteins. *Hum. Mol. Genet.* **11:** 191–198

Brown CE, Lechner T, Howe L & Workman JL (2000) The many HATs of transcription coactivators. *Trends Biochem. Sci.* **25:** 15–19

Brown JM, Green J, das Neves RP, Wallace HAC, Smith AJH, Hughes J, Gray N, Taylor S, Wood WG, Higgs DR, Iborra FJ & Buckle VJ (2008) Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *J. Cell Biol.* **182:** 1083–1097

Brown JM, Leach J, Reittie JE, Atzberger A, Lee-Prudhoe J, Wood WG, Higgs DR, Iborra FJ & Buckle VJ (2006) Coregulated human globin genes are frequently in spatial proximity when active. *J. Cell Biol.* **172:** 177–187

Bulteau A-L, O'Neill HA, Kennedy MC, Ikeda-Saito M, Isaya G & Szweda LI (2004) Frataxin acts as an iron chaperone protein to modulate mitochondrial aconitase activity. *Science* **305:** 242–245

Burnett R, Melander C, Puckett JW, Son LS, Wells RD, Dervan PB & Gottesfeld JM (2006) DNA sequence-specific polyamides alleviate transcription inhibition associated with long GAA·TTC repeats in Friedreich's ataxia. *Proc. Natl. Acad. Sci. U. S. A.* **103:** 11497–11502

Buxbaum AR, Haimovich G & Singer RH (2015) In the right place at the right time: visualizing and understanding mRNA localization. *Nat. Rev. Mol. Cell Biol.* **16:** 95–109

Campuzano V, Montermini L, Lutz Y, Cova L, Hindelang C, Jiralerspong S, Trottier Y, Kish SJ, Faucheux B, Trouillas P, Authier FJ, Dürr A, Mandel J-L, Vescovi A, Pandolfo M & Koenig M (1997) Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes. *Hum. Mol. Genet.* **6:** 1771–1780

Campuzano V, Montermini L, Moltò MD, Pianese L, Cossée M, Cavalcanti F, Monros E, Rodius F, Duclos F, Monticelli A, Zara F, Cañizares J, Koutnikova H, Bidichandani SI, Gellera C, Brice A, Trouillas P, De Michele G, Filla A, De Frutos R, et al (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* **271:** 1423–1427

Carey MF, Peterson CL & Smale ST (2009) Chromatin immunoprecipitation (ChIP). *Cold Spring Harb. Protoc.* **2009:** pdb.prot5279

Castaldo I, Pinelli M, Monticelli A, Acquaviva F, Giacchetti M, Filla A, Sacchetti S, Keller S, Avvedimento VE, Chiariotti L & Cocozza S (2008) DNA methylation in intron 1 of the frataxin gene is related to GAA repeat length and age of onset in Friedreich ataxia patients. *J. Med. Genet.* **45:** 808–812

Castel AL, Cleary JD & Pearson CE (2010) Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat. Rev. Mol. Cell Biol.* **11:** 165–170

Cavadini P, Adamec J, Taroni F, Gakh O & Isaya G (2000) Two-step processing of human frataxin by mitochondrial processing peptidase: precursor and intermediate forms are cleaved at different rates. *J. Biol. Chem.* **275:** 41469–41475

Cedar H & Bergman Y (2009) Linking DNA methylation and histone modification: patterns and paradigms. *Nat. Rev. Genet.* **10:** 295–304

Chambeyron S & Bickmore WA (2004) Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev.* **18:** 1119–1130

Chan PK, Torres R, Yandim C, Law PP, Khadayate S, Mauri M, Grosan C, Chapman-Rothe N, Giunti P, Pook M & Festenstein R (2013) Heterochromatinization induced by GAA-repeat hyperexpansion in Friedreich's ataxia can be reduced upon HDAC inhibition by vitamin B3. *Hum. Mol. Genet.* **22:** 2662–2675

Chantrel-Groussard K, Geromel V, Puccio H, Koenig M, Munnich A, Rötig A & Rustin P (2001) Disabled early recruitment of antioxidant defenses in Friedreich's ataxia. *Hum. Mol. Genet.* **10:** 2061–2067

Chapdelaine P, Coulombe Z, Chikh A, Gerard C & Tremblay JP (2013) A potential new therapeutic approach for Friedreich ataxia: induction of frataxin expression with TALE proteins. *Mol. Ther. Nucleic Acids* **2:** e119

Chartrand P, Bertrand E, Singer RH & Long RM (2000) Long, sensitive and high-resolution detection of RNA in situ. *Methods Enzymol.* **318:** 493–506

Chen I-C, Lin H-Y, Lee G-C, Kao S-H, Chen C-M, Wu Y-R, Hsieh-Li H-M, Su M-T & Lee-Chen G-J (2009) Spinocerebellar ataxia type 8 larger triplet expansion alters histone modification and induces RNA foci. *BMC Mol. Biol.* **10:** 9

Chiurazzi P, Grazia Pomponi M, Pietrobono R, Bakker CE, Neri G & Oostra BA (1999) Synergistic effect of histone hyperacetylation and DNA demethylation in the reactivation of the FMR1 gene. *Hum. Mol. Genet.* **8:** 2317–2323

Chiurazzi P, Pomponi MG, Willemsen R, Oostra BA & Neri G (1998) In vitro reactivation of the FMR1 gene involved in Fragile X syndrome. *Hum. Mol. Genet.* **7:** 109–113

Cho DH, Thienes CP, Mahoney SE, Analau E, Filippova GN & Tapscott SJ (2005) Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. *Mol. Cell* **20:** 483–489

Chubb JR, Boyle S, Perry P & Bickmore WA (2002) Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr. Biol.* **12:** 439–445

Chubb JR, Trcek T, Shenoy SM & Singer RH (2006) Transcriptional pulsing of a developmental gene. *Curr. Biol.* **16:** 1018–1025

Chutake YK, Costello WN, Lam C & Bidichandani SI (2014) Altered nucleosome positioning at the transcription start site and deficient transcriptional initiation in Friedreich ataxia. *J. Biol. Chem.* **289:** 15194–15202

Clark RM, De Biase I, Malykhina AP, Al-Mahdawi S, Pook M & Bidichandani SI (2007) The GAA triplet-repeat is unstable in the context of the human FXN locus and displays age-dependent expansions in cerebellum and DRG in a transgenic mouse model. *Hum. Genet.* **120:** 633–640

Clark RM, Dalgliesh GL, Endres D, Gomez M, Taylor J & Bidichandani SI (2004) Expansion of GAA triplet repeats in the human genome: unique origin of the FRDA mutation at the center of an Alu. *Genomics* **83:** 373–383

Colak D, Zaninovic N, Cohen MS, Rosenwaks Z, Yang W-Y, Gerhardt J, Disney MD & Jaffrey SR (2014) Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. *Science* **343:** 1002–1005

Colin F, Martelli A, Clémancey M, Latour J-M, Gambarelli S, Zeppieri L, Birck C, Page A, Puccio H & Ollagnier de Choudens S (2013) Mammalian frataxin controls sulfur production and iron entry during de novo Fe4S4 cluster assembly. *J. Am. Chem. Soc.* **135:** 733–740

Colombo R & Carobene A (2000) Age of the intronic GAA triplet repeat expansion mutation in Friedreich ataxia. *Hum. Genet.* **106:** 455–458

Condò I, Ventura N, Malisan F, Rufini A, Tomassini B & Testi R (2007) In vivo maturation of human frataxin. *Hum. Mol. Genet.* **16:** 1534–1540

Cook JD, Bencze KZ, Jankovic AD, Crater AK, Busch CN, Bradley PB, Stemmler AJ, Spaller MR & Stemmler TL (2006) Monomeric yeast frataxin is an iron binding protein. *Biochemistry* **45:** 7767–7777

Cossée M, Dürr A, Schmitt M, Dahl N, Trouillas P, Allinson P, Kostrzewa M, Nivelon-Chevallier A, Gustavson K-H, Kohlschütter A, Müller U, Mandel J-L, Brice A, Koenig M, Cavalcanti F, Tammaro A, De Michele G, Filla A, Cocozza S, Labuda M, et al (1999) Friedreich's ataxia: point mutations and clinical presentation of compound heterozygotes. *Ann. Neurol.* **45:** 200–206

Cossée M, Puccio H, Gansmuller A, Koutnikova H, Dierich A, LeMeur M, Fischbeck K, Dollé P & Kœnig M (2000) Inactivation of the Friedreich ataxia mouse gene leads to early embryonic lethality without iron accumulation. *Hum. Mol. Genet.* **9:** 1219–1226

Cossée M, Schmitt M, Campuzano V, Reutenauer L, Moutou C, Mandel J-L & Koenig M (1997) Evolution of the Friedreich's ataxia trinucleotide repeat expansion: Founder effect and premutations. *Proc. Natl. Acad. Sci. U. S. A.* **94:** 7452–7457

Coulon A, Chow CC, Singer RH & Larson DR (2013) Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nat. Rev. Genet.* **14:** 572–584

Di Croce L & Helin K (2013) Transcriptional regulation by Polycomb group proteins. *Nat. Struct. Mol. Biol.* **20:** 1147–1155

Croft JA, Bridger JM, Boyle S, Perry P, Teague P & Bickmore WA (1999) Differences in the localization and morphology of chromosomes in the human nucleus. *J. Cell Biol.* **145:** 1119–1131

Darzacq X, Shav-Tal Y, de Turris V, Brody Y, Shenoy SM, Phair RD & Singer RH (2007) In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **14:** 796–806

Daughters RS, Tuttle DL, Gao W, Ikeda Y, Moseley ML, Ebner TJ, Swanson MS & Ranum LPW (2009) RNA gain-of-function in spinocerebellar ataxia type 8. *PLoS Genet.* **5:** e1000600

Deininger P (2011) Alu elements: know the SINEs. *Genome Biol.* **12:** 236

DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, Nicholson AM, Finch NA, Flynn H, Adamson J, Kouri N, Wojtas A, Sengdy P, Hsiung G-YR, Karydas A, Seeley WW, Josephs KA, Coppola G, Geschwind DH, Wszolek ZK, et al (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72:** 245–256

Dillon N (2008) The impact of gene location in the nucleus on transcriptional regulation. *Dev. Cell* **15:** 182–186

Dion V & Wilson JH (2009) Instability and chromatin structure of expanded trinucleotide repeats. *Trends Genet.* **25:** 288–297

Ditch S, Sammarco MC, Banerjee A & Grabczyk E (2009) Progressive GAA·TTC repeat expansion in human cell lines. *PLoS Genet.* **5:** e1000704

Du J, Campau E, Soragni E, Ku S, Puckett JW, Dervan PB & Gottesfeld JM (2012) Role of mismatch repair enzymes in GAA·TTC triplet-repeat expansion in Friedreich ataxia induced pluripotent stem cells. *J. Biol. Chem.* **287:** 29861–29872

Dürr A, Cossee M, Agid Y, Campuzano V, Mignard C, Penet C, Mandel J-L, Brice A & Koenig M (1996) Clinical and genetic abnormalities in patients with Friedreich's ataxia. *N. Engl. J. Med.* **335:** 1169–1175

Dundr M (2012) Nuclear bodies: multifunctional companions of the genome. *Curr. Opin. Cell Biol.* **24:** 415–422

Dundr M & Misteli T (2003) Measuring dynamics of nuclear proteins by photobleaching. *Curr. Protoc. cell Biol.* **18:** 13.5.1–13.5.18

Eckert D, Buhl S, Weber S, Jäger R & Schorle H (2005) The AP-2 family of transcription factors. *Genome Biol.* **6:** 246

Englander EW & Howard BH (1995) Nucleosome positioning by human Alu elements in chromatin. *J. Biol. Chem.* **270:** 10091–10096

Epplen C, Epplen JT, Frank G, Miterski B, Santos EJM & Schöls L (1997) Differential stability of the (GAA)n tract in the Friedreich ataxia (STM7) gene. *Hum. Genet.* **99:** 834–836

Evans-Galea M V, Carrodus N, Rowley SM, Corben LA, Tai G, Saffery R, Galati JC, Wong NC, Craig JM, Lynch DR, Regner SR, Brocht AFD, Perlman SL, Bushara KO, Gomez CM, Wilmot GR, Li L, Varley E, Delatycki MB & Sarsero JP (2012) FXN

methylation predicts expression and clinical outcome in Friedreich ataxia. *Ann. Neurol.* **71:** 487–497

Evans-Galea M V, Hannan AJ, Carrodus N, Delatycki MB & Saffery R (2013) Epigenetic modifications in trinucleotide repeat diseases. *Trends Mol. Med.* **19:** 655–663

Evans-Galea M V, Pébay A, Dottori M, Corben LA, Ong SH, Lockhart PJ & Delatycki MB (2014) Cell and gene therapy for Friedreich ataxia: progress to date. *Hum. Gene Ther.* **25:** 684–693

Ezzatizadeh V, Pinto RM, Sandi C, Sandi M, Al-Mahdawi S, te Riele H & Pook MA (2012) The mismatch repair system protects against intergenerational GAA repeat instability in a Friedreich ataxia mouse model. *Neurobiol. Dis.* **46:** 165–171

Ezzatizadeh V, Sandi C, Sandi M, Anjomani-Virmouni S, Al-Mahdawi S & Pook MA (2014) MutLα heterodimers modify the molecular phenotype of Friedreich ataxia. *PLoS One* **9:** e100523

Femino a. M (1998) Visualization of single RNA transcripts in situ. *Science* **280:** 585–590

Ferrai C, de Castro IJ, Lavitas L, Chotalia M & Pombo A (2010a) Gene positioning. *Cold Spring Harb. Perspect. Biol.* **2:** a000588

Ferrai C, Xie SQ, Luraghi P, Munari D, Ramirez F, Branco MR, Pombo A & Crippa MP (2010b) Poised transcription factories prime silent uPA gene prior to activation. *PLoS Biol.* **8:** e1000270

Ferrante RJ, Kubilus JK, Lee J, Ryu H, Beesen A, Zucker B, Smith K, Kowall NW, Ratan RR, Luthi-Carter R & Hersch SM (2003) Histone deacetylase inhibition by sodium butyrate chemotherapy ameliorates the neurodegenerative phenotype in Huntington's disease mice. *J. Neurosci.* **23:** 9418–9427

Filla A, DeMichele G, Caruso G, Marconi R & Campanella G (1990) Genetic data and natural history of Friedreich's disease: a study of 80 Italian patients. *J. Neurol.* **237:** 345–351

Filla A, De Michele G, Cavalcanti F, Pianese L, Monticelli A, Campanella G & Cocozza S (1996) The relationship between trinucleotide (GAA) repeat length and clinical features in Friedreich ataxia. *Am. J. Hum. Genet.* **59:** 554–560

Finlan LE, Sproul D, Thomson I, Boyle S, Kerr E, Perry P, Ylstra B, Chubb JR & Bickmore W a (2008) Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet.* **4:** e1000039

Fleming J, Spinoulas A, Zheng M, Cunningham SC, Ginn SL, McQuilty RC, Rowe PB & Alexander IE (2005) Partial correction of sensitivity to oxidant stress in Friedreich ataxia patient fibroblasts by frataxin-encoding adeno-associated virus and lentivirus vectors. *Hum. Gene Ther.* **16:** 947–956

Fox AH, Lam YW, Leung AKL, Lyon CE, Andersen J, Mann M & Lamond AI (2002) Paraspeckles: a novel nuclear domain. *Curr. Biol.* **12:** 13–25

Friedreich N (1863a) Ueber degenerative Atrophie der spinalen Hinterstränge (On degenerative atrophy of the spinal dorsal columns). *Virchows Arch. A. Pathol. Pathol. Anat.* **26:** 391–419

Friedreich N (1863b) Ueber degenerative Atrophie der spinalen Hinterstränge (On degenerative atrophy of the spinal dorsal columns). *Virchows Arch. A. Pathol. Pathol. Anat.* **26:** 433–459

Friedreich N (1863c) Ueber degenerative Atrophie der spinalen Hinterstränge (On degenerative atrophy of the spinal dorsal columns). *Virchows Arch. A. Pathol. Pathol. Anat.* **27:** 1–26

Friedreich N (1876) Ueber Ataxie mit besonderer Berücksichtigung der hereditären Formen (About ataxia with special consideration of the hereditary forms). *Virchows Arch. A. Pathol. Pathol. Anat.* **68:** 145–245

Friedreich N (1877) Ueber Ataxie mit besonderer Berücksichtigung der hereditären Formen. Nachtrag (About ataxia with special consideration of the hereditary forms. Postscriptum). *Virchows Arch. A. Pathol. Pathol. Anat.* **70:** 140–152

Fry M & Loeb LA (1994) The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahelical structure. *Proc. Natl. Acad. Sci. U. S. A.* **91:** 4950–4954

Fusco D, Accornero N, Lavoie B, Shenoy SM, Blanchard JM, Singer RH & Bertrand E (2003) Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. *Curr. Biol.* **13:** 161–167

Gal-Yam EN, Egger G, Iniguez L, Holster H, Einarsson S, Zhang X, Lin JC, Liang G, Jones PA & Tanay A (2008) Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. *Proc. Natl. Acad. Sci. U. S. A.* **105:** 12979–12984

Gatchel JR & Zoghbi HY (2005) Diseases of unstable repeat expansion: mechanisms and common principles. *Nat. Rev. Genet.* **6:** 743–755

Gellera C, Castellotti B, Mariotti C, Mineri R, Seveso V, DiDonato S & Taroni F (2007) Frataxin gene point mutations in Italian Friedreich ataxia patients. *Neurogenetics* **8:** 289–299

Gerber J, Mühlenhoff U & Lill R (2003) An interaction between frataxin and Isu1/Nfs1 that is crucial for Fe/S cluster synthesis on Isu1. *EMBO Rep.* **4:** 906–911

Gilbert N, Boyle S, Fiegler H, Woodfine K, Carter NP & Bickmore WA (2004) Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell* **118:** 555–566

Gimenez-Cassina A, Wade-Martins R, Gomez-Sebastian S, Corona J-C, Lim F & Diaz-Nido J (2011) Infectious delivery and long-term persistence of transgene expression in the brain by a 135-kb iBAC-FXN genomic DNA expression vector. *Gene Ther.* **18:** 1015–1019

Golding I, Paulsson J, Zawilski SM & Cox EC (2005) Real-time kinetics of gene activity in individual bacteria. *Cell* **123:** 1025–1036

Gomes-Pereira M, Bidichandani S & Monckton D (2004a) Analysis of unstable triplet repeats using small-pool polymerase chain reaction. *Methods Mol. Biol.* **277:** 61–76

Gomes-Pereira M, Fortune MT, Ingram L, McAbney JP & Monckton DG (2004b) Pms2 is a genetic enhancer of trinucleotide CAG·CTG repeat somatic mosaicism: implications for the mechanism of triplet repeat expansion. *Hum. Mol. Genet.* **13:** 1815–1825

Gomez-Sebastian S, Gimenez-Cassina A, Diaz-Nido J, Lim F & Wade-Martins R (2007) Infectious Delivery and Expression of a 135 kb Human FRDA Genomic DNA Locus Complements Friedreich's Ataxia Deficiency in Human Cells. *Mol. Ther.* **15:** 248–254

González-Cabo P, Vázquez-Manrique RP, García-Gimeno MA, Sanz P & Palau F (2005) Frataxin interacts functionally with mitochondrial electron transport chain proteins. *Hum. Mol. Genet.* **14:** 2091–2098

Gottesfeld JM, Rusche JR & Pandolfo M (2013) Increasing frataxin gene expression with histone deacetylase inhibitors as a therapeutic approach for Friedreich's ataxia. *J. Neurochem.* **126:** 147–154

Grabczyk E, Mancuso M & Sammarco MC (2007) A persistent RNA·DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP in vitro. *Nucleic Acids Res.* **35:** 5351–5359

Grabczyk E & Usdin K (2000) The GAA • TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. *Nucleic Acids Res.* **28:** 2815–2822

Grant L, Sun J, Xu H, Subramony SH, Chaires JB & Hebert MD (2006) Rational selection of small molecules that increase transcription through the GAA repeats found in Friedreich's ataxia. *FEBS Lett.* **580:** 5399–5405

Greene E, Entezam A, Kumari D & Usdin K (2005) Ancient repeated DNA elements and the regulation of the human frataxin promoter. *Genomics* **85:** 221–230

Greene E, Mahishi L, Entezam A, Kumari D & Usdin K (2007) Repeat-induced epigenetic changes in intron 1 of the frataxin gene and its consequences in Friedreich ataxia. *Nucleic Acids Res.* **35:** 3383–3390

Groh M, Lufino MMP, Wade-Martins R & Gromak N (2014) R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS Genet.* **10:** e1004318

Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W & van Steensel B (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453:** 948–951

Guenther MG, Levine SS, Boyer L a, Jaenisch R & Young R a (2007) A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130:** 77–88

Haeusler AR, Donnelly CJ, Periz G, Simko EAJ, Shaw PG, Kim M-S, Maragakis NJ, Troncoso JC, Pandey A, Sattler R, Rothstein JD & Wang J (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507:** 195–200

Haim-Vilmovsky L, Gadir N, Herbst RH & Gerst JE (2011) A genomic integration method for the simultaneous visualization of endogenous mRNAs and their translation products in living yeast. *RNA* **17:** 2249–2255

Halabi A, Ditch S, Wang J & Grabczyk E (2012) DNA mismatch repair complex MutSbeta promotes GAA·TTC repeat expansion in human cells. *J. Biol. Chem.* **287:** 29958–29967

Harrison PR, Conkie D, Paul J & Jones K (1973) Localisation of cellular globin messenger RNA by in situ hybridisation to complementary DNA. *FEBS Lett.* **32:** 109–112

Herman D, Jenssen K, Burnett R, Soragni E, Perlman SL & Gottesfeld JM (2006) Histone deacetylase inhibitors reverse gene silencing in Friedreich's ataxia. *Nat. Chem. Biol.* **2:** 551–558

Hick A, Wattenhofer-Donzé M, Chintawar S, Tropel P, Simard JP, Vaucamps N, Gall D, Lambot L, André C, Reutenauer L, Rai M, Teletin M, Messaddeq N, Schiffmann SN, Viville S, Pearson CE, Pandolfo M & Puccio HM (2013) Neurons and

cardiomyocytes derived from induced pluripotent stem cells as a model for mitochondrial defects in Friedreich's ataxia. *Dis. Model. Mech.* **6:** 608–621

Hocine S, Raymond P, Zenklusen D, Chao J a & Singer RH (2013) Single-molecule analysis of gene expression using two-color RNA labeling in live yeast. *Nat. Methods* **10:** 119–121

Holloway TP, Rowley SM, Delatycki MB & Sarsero JP (2011) Detection of interruptions in the GAA trinucleotide repeat expansion in the FXN gene of Friedreich ataxia. *Biotechniques* **50:** 182–186

Holt I, Mittal S, Furling D, Butler-Browne GS, David Brook J & Morris GE (2007) Defective mRNA in myotonic dystrophy accumulates at the periphery of nuclear splicing speckles. *Genes to Cells* **12:** 1035–1048

Huynen MA, Snel B, Bork P & Gibson TJ (2001) The phylogenetic distribution of frataxin indicates a role in iron-sulfur cluster protein assembly. *Hum. Mol. Genet.* **10:** 2463–2468

Iannuccelli E, Mompart F, Gellin J, Lahbib-Mansais Y, Yerle M & Boudier T (2010) NEMO: a tool for analyzing gene and chromosome territory distributions from 3D-FISH experiments. *Bioinformatics* **26:** 696–697

Iborra FJ, Pombo A, Jackson DA & Cook PR (1996) Active RNA polymerases are localized within discrete transcription 'factories' in human nuclei. *J. Cell Sci.* **109:** 1427–1436

Imbert G, Kretz C, Johnson K & Mandel J-L (1993) Origin of the expansion mutation in myotonic dystrophy. *Nat. Genet.* **4:** 72–76

Iyer RR, Pluciennik A, Napierala M & Wells RD (2015) DNA triplet repeat expansion and mismatch repair. *Annu. Rev. Biochem.* **84:** 199–226

Jakovcevski M & Akbarian S (2012) Epigenetic mechanisms in neurological disease. *Nat. Med.* **18:** 1194–1204

Janicki SM, Tsukamoto T, Salghetti SE, Tansey WP, Sachidanandam R, Prasanth K V, Ried T, Shav-Tal Y, Bertrand E, Singer RH & Spector DL (2004) From silencing to gene expression: real-time analysis in single cells. *Cell* **116:** 683–698

Jenuwein T & Allis CD (2001) Translating the histone code. *Science* **293:** 1074–1080

Jia H, Pallos J, Jacques V, Lau A, Tang B, Cooper A, Syed A, Purcell J, Chen Y, Sharma S, Sangrey GR, Darnell SB, Plasterer H, Sadri-Vakili G, Gottesfeld JM, Thompson LM, Rusche JR, Marsh JL & Thomas EA (2012) Histone deacetylase (HDAC) inhibitors targeting HDAC3 and HDAC1 ameliorate polyglutamine-elicited phenotypes in model systems of Huntington's disease. *Neurobiol. Dis.* **46:** 351–361

Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13:** 484–492

Juvonen V, Kulmala S-M, Ignatius J, Penttinen M & Savontaus M-L (2002) Dissecting the epidemiology of a trinucleotide repeat disease – example of FRDA in Finland. *Hum. Genet.* **110:** 36–40

Katz ZB, Wells AL, Park HY, Wu B, Shenoy SM & Singer RH (2012) β-Actin mRNA compartmentalization enhances focal adhesion stability and directs cell migration. *Genes Dev.* **26:** 1885–1890

Kelly TK, Miranda TB, Liang G, Berman BP, Lin JC, Tanay A & Jones PA (2010) H2A.Z maintenance during mitosis reveals nucleosome shifting on mitotically silenced genes. *Mol. Cell* **39:** 901–911

Kim E, Napierala M & Dent SYR (2011) Hyperexpansion of GAA repeats affects post-initiation steps of FXN transcription in Friedreich's ataxia. *Nucleic Acids Res.* **39:** 8366–8377

Kind J, Pagie L, Ortabozkoyun H, Boyle S, de Vries SS, Janssen H, Amendola M, Nolen LD, Bickmore W a & van Steensel B (2013) Single-cell dynamics of genome-nuclear lamina interactions. *Cell* **153:** 178–192

Kind J & van Steensel B (2010) Genome-nuclear lamina interactions and gene regulation. *Curr. Opin. Cell Biol.* **22:** 320–325

Klesert T, Otten A, Bird T & Tapscott S (1997) Trinucleotide repeat expansion at the myotonic dystrophy locus reduces expression of DMAHP. *Nat. Genet.* **16:** 402–406

Knöll B & Nordheim A (2009) Functional versatility of transcription factors in the nervous system: the SRF paradigm. *Trends Neurosci.* **32:** 432–442

Koeppen A, Michael S, Knutson M, Haile D, Qian J, Levi S, Santambrogio P, Garrick M & Lamarche J (2007) The dentate nucleus in Friedreich's ataxia: the role of iron-responsive proteins. *Acta Neuropathol.* **114:** 163–173

Koeppen A, Ramirez RL, Yu D, Collins S, Qian J, Parsons P, Yang K, Chen Z, Mazurkiewicz J & Feustel P (2012) Friedreich's ataxia causes redistribution of iron, copper, and zinc in the dentate nucleus. *The Cerebellum* **11:** 845–860

Koeppen AH, Kuntzsch EC, Bjork ST, Ramirez RL, Mazurkiewicz JE & Feustel PJ (2013) Friedreich ataxia: metal dysmetabolism in dorsal root ganglia. *Acta Neuropathol. Commun.* **1:** 26

Kornberg RD & Thonmas JO (1974) Chromatin structure: oligomers of the histones. *Science* **184:** 865–868

Kosak ST, Skok JA, Medina KL, Riblet R, Le Beau MM, Fisher AG & Singh H (2002) Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* **296:** 158–162

Koutnikova H, Campuzano V, Foury F, Dolle P, Cazzalini O & Koenig M (1997) Studies of human, mouse and yeast homologues indicate a mitochondrial function for frataxin. *Nat. Genet.* **16:** 345–351

Koutnikova H, Campuzano V & Koenig M (1998) Maturation of wild-type and mutated frataxin by the mitochondrial processing peptidase. *Hum. Mol. Genet.* **7:** 1485–1489

Kouzarides T (2007) Chromatin modifications and their function. *Cell* **128:** 693–705

Ku S, Soragni E, Campau E, Thomas EA, Altun G, Laurent LC, Loring JF, Napierala M & Joel Mg (2010) Friedreich's ataxia induced pluripotent stem cells model intergenerational GAA•TTC triplet repeat instability. *Cell Stem Cell* **7:** 631–637

Kumaran RI & Spector DL (2008) A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence. *J. Cell Biol.* **180:** 51–65

Kumari D, Biacsi RE & Usdin K (2011) Repeat expansion affects both transcription initiation and elongation in friedreich ataxia cells. *J. Biol. Chem.* **286:** 4209–4215

Kumari D & Usdin K (2010) The distribution of repressive histone modifications on silenced FMR1 alleles provides clues to the mechanism of gene silencing in fragile X syndrome. *Hum. Mol. Genet.* **19:** 4634–4642

Kumari D & Usdin K (2012) Is Friedreich ataxia an epigenetic disorder? *Clin. Epigenetics* **4:** 1–10

Kumari D & Usdin K (2014) Polycomb group complexes are recruited to reactivated FMR1 alleles in Fragile X syndrome in response to FMR1 transcription. *Hum. Mol. Genet.* **23:** 6575–6583

Labuda M, Labuda D, Miranda C, Poirier J, Soong B-W, Barucha NE & Pandolfo M (2000) Unique origin and specific ethnic distribution of the Friedreich ataxia GAA expansion. *Neurology* **54:** 2322–2324

Lamarche JB, Côté M & Lemieux B (1980) The cardiomyopathy of Friedreich's ataxia morphological observations in 3 cases. *Can. J. Neurol. Sci.* **7:** 389–396

Lange S, Katayama Y, Schmid M, Burkacky O, Bräuchle C, Lamb DC & Jansen R-P (2008) Simultaneous transport of different localized mRNA species revealed by live-cell imaging. *Traffic* **9:** 1256–1267

Levesque MJ & Raj A (2013) M. J. Levesque, A. Raj. (University of Pennsylvania).

Levsky JM & Singer RH (2003) Gene expression and the myth of the average cell. *Trends Cell Biol.* **13:** 4–6

Li B, Carey M & Workman JL (2007) The role of chromatin during transcription. *Cell* **128:** 707–719

Li G-M (2008) Mechanisms and functions of DNA mismatch repair. *Cell Res* **18:** 85–98

Li K, Besse EK, Ha D, Kovtunovych G & Rouault TA (2008) Iron-dependent regulation of frataxin expression: implications for treatment of Friedreich ataxia. *Hum. Mol. Genet.* **17:** 2265–2273

Li K, Singh A, Crooks DR, Dai X, Cong Z, Pan L, Ha D & Rouault TA (2010) Expression of human frataxin is regulated by transcription factors SRF and TFAP2. *PLoS One* **5:** e12286

Li L, Voullaire L, Sandi C, Pook MA, Ioannou PA, Delatycki MB & Sarsero JP (2013) Pharmacological screening using an FXN-EGFP cellular genomic reporter assay for the therapy of Friedreich ataxia. *PLoS One* **8:** e55940

Li Y, Polak U, Bhalla AD, Rozwadowska N, Butler JS, Lynch DR, Dent SYR & Napierala M (2015) Excision of expanded GAA repeats alleviates the molecular phenotype of Friedreich's Ataxia. *Mol. Ther.* **23:** 1055–1065

Libri V, Yandim C, Athanasopoulos S, Loyse N, Natisvili T, Law PP, Chan PK, Mohammad T, Mauri M, Tam KT, Leiper J, Piper S, Ramesh A, Parkinson MH, Huson L, Giunti P & Festenstein R (2014) Epigenetic and neurological effects and safety of high-dose nicotinamide in patients with Friedreich's ataxia: an exploratory, open-label, dose-escalation study. *Lancet* **384:** 504–513

Lin Y, Dion V & Wilson JH (2006) Transcription promotes contraction of CAG repeat tracts in human cells. *Nat. Struct. Mol. Biol.* **13:** 179–180

Lionnet T, Czaplinski K, Darzacq X, Shav-Tal Y, Wells AL, Chao J a, Park HY, de Turris V, Lopez-Jones M & Singer RH (2011) A transgenic mouse for in vivo detection of endogenous labeled mRNA. *Nat. Methods* **8:** 165–170

Liu G, Chen X, Bissler JJ, Sinden RR & Leffak M (2010) Replication dependent instability at (CTG)•(CAG) repeat hairpins in human cells. *Nat. Chem. Biol.* **6:** 652–659

Liu J, Verma P, Evans-Galea M, Delatycki M, Michalska A, Leung J, Crombie D, Sarsero J, Williamson R, Dottori M & Pébay A (2011) Generation of induced pluripotent stem cell lines from Friedreich ataxia patients. *Stem Cell Rev. Reports* **7:** 703–713

Lockhart DJ & Winzeler EA (2000) Genomics, gene expression and DNA arrays. *Nature* **405:** 827–836

Lokanga RA, Zhao X-N & Usdin K (2014) The mismatch repair protein MSH2 is rate limiting for repeat expansion in a Fragile X premutation mouse model. *Hum. Mutat.* **35:** 129–136

Long RM, Gu W, Lorimer E, Singer RH & Chartrand P (2000) She2p is a novel RNA-binding protein that recruits the Myo4p–She3p complex to ASH1 mRNA. *EMBO J.* **19:** 6592–6601

López Castel A, Nakamori M, Tomé S, Chitayat D, Gourdon G, Thornton CA & Pearson CE (2011) Expanded CTG repeat demarcates a boundary for abnormal CpG methylation in myotonic dystrophy patient tissues. *Hum. Mol. Genet.* **20:** 1–15

Lufino MMP, Manservigi R & Wade-Martins R (2007) An S/MAR-based infectious episomal genomic DNA expression vector provides long-term regulated functional complementation of LDLR deficiency. *Nucleic Acids Res.* **35:** e98

Lufino MMP, Silva AM, Németh AH, Alegre-Abarrategui J, Russell AJ & Wade-Martins R (2013) A GAA repeat expansion reporter model of Friedreich's ataxia recapitulates the genomic context and allows rapid screening of therapeutic compounds. *Hum. Mol. Genet.* **22:** 5173–5187

Malter HE, Iber JC, Willemsen R, Graaff E de, Tarleton JC, Leisti J, Warren ST & Oostra BA (1997) Characterization of the full fragile X syndrome mutation in fetal gametes. *Nat. Genet.* **15:** 165–169

Manley K, Shirley TL, Flaherty L & Messer A (1999) Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat. Genet.* **23:** 471–473

Mariappan SVS, Catasti P, Silks III LA, Bradbury EM & Gupta G (1999) The high-resolution structure of the triplex formed by the GAA/TTC triplet repeat associated with Friedreich's ataxia1. *J. Mol. Biol.* **285:** 2035–2052

Marquis Gacy A, Goellner G, Juranić N, Macura S & McMurray CT (1995) Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell* **81:** 533–540

Martelli A, Napierala M & Puccio H (2012) Understanding the genetic and molecular pathogenesis of Friedreich's ataxia through animal and cellular models. *Dis. Model. Mech.* **5:** 165–176

Martelli A & Puccio H (2014) Dysregulation of cellular iron metabolism in Friedreich ataxia: from primary iron-sulfur cluster deficit to mitochondrial iron accumulation. *Front. Pharmacol.* **5:** 130

McMurray CT (2010) Mechanisms of trinucleotide repeat instability during human development. *Nat. Rev. Genet.* **11:** 786–799

McStay B & Grummt I (2008) The epigenetics of rRNA genes: from molecular to chromosome biology. *Annu. Rev. Cell Dev. Biol.* **24:** 131–157

Meaburn KJ & Misteli T (2008) Locus-specific and activity-independent gene repositioning during early tumorigenesis. *J. Cell Biol.* **180:** 39–50

Meuleman W, Peric-Hupkes D, Kind J, Beaudry J-B, Pagie L, Kellis M, Reinders M, Wessels L & van Steensel B (2013) Constitutive nuclear lamina–genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res.* **23:** 270–280

Michael S, Petrocine S, Qian J, Lamarche J, Knutson M, Garrick M & Koeppen A (2006) Iron and iron-responsive proteins in the cardiomyopathy of Friedreich's ataxia. *The Cerebellum* **5:** 257–267

Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, Wingett SW, Andrews S, Grey W, Ewels PA, Herman B, Happe S, Higgs A, LeProust E,

Follows GA, Fraser P, Luscombe NM & Osborne CS (2015) Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47:** 598–606

Mirkin SM (2007) Expandable DNA repeats and human disease. *Nature* **447:** 932–940

Misteli T (2007) Beyond the sequence: cellular organization of genome function. *Cell* **128:** 787–800

Monrós E, Moltó MD, Martínez F, Canizares J, Blanca J, Vílchez JJ, Prieto F, Frutos R de & Palau F (1997) Phenotype correlation and intergenerational dynamics of the Friedreich ataxia GAA trinucleotide repeat. *Am. J. Hum. Genet.* **61:** 101–110

Montermini L, Andermann E, Labuda M, Richter A, Pandolfo M, Cavalcanti F, Pianese L, Iodice L, Farina G, Monticelli A, Turano M, Filla A, De Michele G & Cocozza S (1997a) The Friedreich ataxia GAA triplet repeat: premutation and normal alleles. *Hum. Mol. Genet.* **6:** 1261–1266

Montermini L, Richeter A, Morgan K, Justice CM, Julien D, Castellotti B, Mercier J, Poirier J, Capozzoli F, Bouchard J-P, Lemieux B, Mathieu J, Vanasse M, Seni M-H, Graham G, Andermann F, Andermann E, Melançon SB, Keats BJB, Di Donato S, et al (1997b) Phenotypic variability in friedreich ataxia: Role of the associated GAA triplet repeat expansion. *Ann. Neurol.* **41:** 675–682

Monticelli A, Giacchetti M, De Biase I, Pianese L, Turano M, Pandolfo M & Cocozza S (2004) New clues on the origin of the Friedreich ataxia expanded alleles from the analysis of new polymorphisms closely linked to the mutation. *Hum. Genet.* **114:** 458–463

Mori K, Weng S-M, Arzberger T, May S, Rentzsch K, Kremmer E, Schmid B, Kretzschmar HA, Cruts M, Van Broeckhoven C, Haass C & Edbauer D (2013) The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science* **339:** 1335–1338

Moseley ML, Zu T, Ikeda Y, Gao W, Mosemiller AK, Daughters RS, Chen G, Weatherspoon MR, Clark HB, Ebner TJ, Day JW & Ranum LPW (2006) Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. *Nat. Genet.* **38:** 758–769

Muramoto T, Cannon D, Gierlinski M, Corrigan A, Barton GJ & Chubb JR (2012) Live imaging of nascent RNA dynamics reveals distinct types of transcriptional pulse regulation. *Proc. Natl. Acad. Sci. U. S. A.* **109:** 7350–7355

Nelson DL, Orr HT & Warren ST (2013) The unstable repeats—three evolving faces of neurological disease. *Neuron* **77:** 825–843

Nguyen CT, Gonzales FA & Jones PA (2001) Altered chromatin structure associated with methylation-induced gene silencing in cancer cells: correlation of accessibility, methylation, MeCP2 binding and acetylation. *Nucleic Acids Res.* **29:** 4598–4606

Norris J, Fan D, Aleman C, Marks JR, Futreal PA, Wiseman RW, Iglehart JD, Deininger PL & McDonnell DP (1995) Identification of a new subclass of Alu DNA repeats which can function as estrogen receptor-dependent transcriptional enhancers. *J. Biol. Chem.* **270:** 22777–22782

Novelli G, Spedini G, Destro-Bisol G, Gennarelli M, Fattorini C & Dallapiccola B (1994) North Eurasian origin of the myotonic dystrophy mutation. *Hum. Mutat.* **4:** 79–81

O'Neill HA, Gakh O, Park S, Cui J, Mooney SM, Sampson M, Ferreira GC & Isaya G (2005) Assembly of human Frataxin is a mechanism for detoxifying redox-active Iron. *Biochemistry* **44:** 537–545

Ohshima K, Montermini L, Wells RD & Pandolfo M (1998) Inhibitory effects of expanded GAA·TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication in vivo. *J. Biol. Chem.* **273:** 14588–14595

Oliviero S & Monaci P (1988) RNA polymerase III promoter elements enhance transcription of RNA polymerase II genes. *Nucleic Acids Res.* **16:** 1285–1293

Osborne CS, Chakalova L, Brown KE, Carter D, Horton A, Debrand E, Goyenechea B, Mitchell JA, Lopes S, Reik W & Fraser P (2004) Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.* **36:** 1065–1071

Otten AD & Tapscott SJ (1995) Triplet repeat expansion in myotonic dystrophy alters the adjacent chromatin structure. *Proc. Natl. Acad. Sci. U. S. A.* **92:** 5465–5469

Owen BAL, Yang Z, Lai M, Gajek M, Badger JD, Hayes JJ, Edelmann W, Kucherlapati R, Wilson TM & McMurray CT (2005) (CAG)n-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. *Nat. Struct. Mol. Biol.* **12:** 663–670

Padeken J & Heun P (2014) Nucleolus and nuclear periphery: velcro for heterochromatin. *Curr. Opin. Cell Biol.* **28:** 54–60

Pandolfo M (2008) Drug Insight: antioxidant therapy in inherited ataxias. *Nat Clin Pr. Neuro* **4:** 86–96

Pandolfo M (2009) Friedreich ataxia: The clinical picture. *J. Neurol.* **256:** 3–8

Pandolfo M & Hausmann L (2013) Deferiprone for the treatment of Friedreich's ataxia. *J. Neurochem.* **126:** 142–146

Park HY, Lim H, Yoon YJ, Follenzi A, Nwokafor C, Lopez-Jones M, Meng X & Singer RH (2014) Visualization of dynamics of single endogenous mRNA labeled in live mouse. *Science* **343:** 422–424

Parkinson MH, Schulz JB & Giunti P (2013) Co-enzyme Q10 and idebenone use in Friedreich's ataxia. *J. Neurochem.* **126:** 125–141

Pastore A & Puccio H (2013) Frataxin: a protein in search for a function. *J. Neurochem.* **126:** 43–52

Perdomini M, Belbellaa B, Monassier L, Reutenauer L, Messaddeq N, Cartier N, Crystal RG, Aubourg P & Puccio H (2014) Prevention and reversal of severe mitochondrial cardiomyopathy by gene therapy in a mouse model of Friedreich's ataxia. *Nat. Med.* **20:** 542–547

Perdomini M, Hick A, Puccio H & Pook MA (2013) Animal and cellular models of Friedreich ataxia. *J. Neurochem.* **126:** 65–79

Peric-Hupkes D, Meuleman W, Pagie L, Bruggeman SWM, Solovei I, Brugman W, Gräf S, Flicek P, Kerkhoven RM, van Lohuizen M, Reinders M, Wessels L & van Steensel B (2010) Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell* **38:** 603–613

Phillips JE & Corces VG (2009) CTCF: master weaver of the genome. *Cell* **137:** 1194–1211

Pianese L, Cavalcanti F, De Michele G, Filla A, Campanella G, Calabrese O, Castaldo I, Monticelli A & Cocozza S (1997) The effect of parental gender on the GAA dynamic mutation in the FRDA gene. *Am. J. Hum. Genet.* **60:** 460–463

Pinto RM, Dragileva E, Kirby A, Lloret A, Lopez E, St. Claire J, Panigrahi GB, Hou C, Holloway K, Gillis T, Guide JR, Cohen PE, Li G-M, Pearson CE, Daly MJ & Wheeler VC (2013) Mismatch Repair Genes Mlh1 and Mlh3 modify CAG instability in Huntington's disease mice: genome-wide and candidate approaches. *PLoS Genet.* **9:** e1003930

Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolsheimer E, Zeitlinger J, Lewitter F, Gifford DK & Young RA (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122:** 517–527

Polak U, McIvor E, Dent SYR, Wells RD & Napierala M (2013) Expanded complexity of unstable repeat diseases. *BioFactors* **39:** 164–175

Polo JM, Calleja J, Combarros O & Berciano J (1991) Hereditary ataxias and paralegias in Cantabria, Spain. *Brain* **114:** 855–866

Pook M, Al-Mahdawi S, Carroll C, Cossée M, Puccio H, Lawrence L, Clark P, Lowrie M, Bradley J, Cooper M, Kœnig M & Chamberlain S (2001) Rescue of the Friedreich's ataxia knockout mouse by human YAC transgenesis. *Neurogenetics* **3:** 185–193

Potaman VN, Oussatcheva EA, Lyubchenko YL, Shlyakhtenko LS, Bidichandani SI, Ashizawa T & Sinden RR (2004) Length-dependent structure formation in Friedreich ataxia (GAA)n·(TTC)n repeats at neutral pH. *Nucleic Acids Res.* **32:** 1224–1231

Prasanth K V, Prasanth SG, Xuan Z, Hearn S, Freier SM, Bennett CF, Zhang MQ & Spector DL (2005) Regulating gene expression through RNA nuclear retention. *Cell* **123:** 249–263

Puccio H, Simon D, Cossee M, Criqui-Filipe P, Tiziano F, Melki J, Hindelang C, Matyas R, Rustin P & Koenig M (2001) Mouse models for Friedreich ataxia exhibit cardiomyopathy, sensory nerve defect and Fe-S enzyme deficiency followed by intramitochondrial iron deposits. *Nat. Genet.* **27:** 181–186

Punga T & Bühler M (2010) Long intronic GAA repeats causing Friedreich ataxia impede transcription elongation. *EMBO Mol. Med.* **2:** 120–129

Qureshi I a & Mehler MF (2010) Impact of nuclear organization and dynamics on epigenetic regulation in the central nervous system: implications for neurological disease states. *Ann. N. Y. Acad. Sci.* **1204:** 20–37

Raap AK, van de Corput MPC, Vervenne RAM, van Gijlswijk RPM, Tanke HJ & Wiegant J (1995) Ultra-sensitive FISH using peroxidase-mediated deposition of biotin- or fluorochrome tyramides. *Hum. Mol. Genet.* **4:** 529–534

Ragoczy T, Bender MA, Telling A, Byron R & Groudine M (2006) The locus control region is required for association of the murine β-globin locus with engaged transcription factories during erythroid maturation. *Genes Dev.* **20:** 1447–1457

Rai M, Soragni E, Chou CJ, Barnes G, Jones S, Rusche JR, Gottesfeld JM & Pandolfo M (2010) Two new pimelic diphenylamide HDAC inhibitors induce sustained frataxin upregulation in cells from Friedreich's ataxia patients and in a mouse model. *PLoS One* **5:** e8825

Rai M, Soragni E, Jenssen K, Burnett R, Herman D, Coppola G, Geschwind DH, Gottesfeld JM & Pandolfo M (2008) HDAC inhibitors correct frataxin deficiency in a Friedreich ataxia mouse model. *PLoS One* **3:** e1958

Raj A, van den Bogaard P, Rifkin S a, van Oudenaarden A & Tyagi S (2008) Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5:** 877–879

Raj A, Peskin CS, Tranchina D, Vargas DY & Tyagi S (2006) Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* **4:** e309

Raj A & Tyagi S (2010) Detection of individual endogenous RNA transcripts in situ using multiple singly labeled probes. *Methods Enzymol.* **472:** 365–386

Ramazzotti A, Vanmansart V & Foury F (2004) Mitochondrial functional interactions between frataxin and Isu1p, the iron–sulfur cluster scaffold protein, in Saccharomyces cerevisiae. *FEBS Lett.* **557:** 215–220

Reddy K, Schmidt MHM, Geist JM, Thakkar NP, Panigrahi GB, Wang Y-H & Pearson CE (2014) Processing of double-R-loops in (CAG)·(CTG) and C9orf72 (GGGGCC)·(GGCCCC) repeats causes instability. *Nucleic Acids Res.* **42:** 10473–10487

Reddy K, Tam M, Bowater RP, Barber M, Tomlinson M, Nichol Edamura K, Wang Y-H & Pearson CE (2011) Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. *Nucleic Acids Res.* **39:** 1749–1762

Reddy KL, Zullo JM, Bertolino E & Singh H (2008) Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* **452:** 243–247

Reinke H & Hörz W (2003) Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. *Mol. Cell* **11:** 1599–1607

Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, Gibbs JR, Schymick JC, Laaksovirta H, van Swieten JC, Myllykangas L, Kalimo H, Paetau A, Abramzon Y, Remes AM, Kaganovich A, Scholz SW, Duckworth J, Ding J, Harmer DW,

Hernandez DG, et al (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72:** 257–268

Ristow M, Pfister MF, Yee AJ, Schubert M, Michael L, Zhang C-Y, Ueki K, Michael MD, Lowell BB & Kahn CR (2000) Frataxin activates mitochondrial energy conversion and oxidative phosphorylation. *Proc. Natl. Acad. Sci.* **97:** 12239–12243

Rodríguez R, Hernández-Hernández O, Magaña JJ, González-Ramírez R, García-López ES & Cisneros B (2015) Altered nuclear structure in myotonic dystrophy type 1-derived fibroblasts. *Mol. Biol. Rep.* **42:** 479–488

Rotig A, de Lonlay P, Chretien D, Foury F, Koenig M, Sidi D, Munnich A & Rustin P (1997) Aconitase and mitochondrial iron-sulphur protein deficiency in Friedreich ataxia. *Nat Genet* **17:** 215–217

Rubinsztein DC, Leggo J, Goodburn S, Barton DE & Ferguson-Smith MA (1995) Haplotype analysis of the Δ2642 and (CAG)n polymorphisms in the Huntington's disease (HD) gene provides an explanation for an apparent 'founder' HD haplotype. *Hum. Mol. Genet.* **4:** 203–206

Sakamoto N, Chastain PD, Parniewski P, Ohshima K, Pandolfo M, Griffith JD & Wells RD (1999) Sticky DNA: self-association properties of long GAA·TTC repeats in R·R·Y triplex structures from Friedreich's ataxia. *Mol. Cell* **3:** 465–475

Sakamoto N, Ohshima K, Montermini L, Pandolfo M & Wells RD (2001) Sticky DNA, a self-associated complex formed at long GAA·TTC repeats in Intron 1 of the Frataxin gene, inhibits transcription. *J. Biol. Chem.* **276:** 27171–27177

Sandi C, Pinto RM, Al-Mahdawi S, Ezzatizadeh V, Barnes G, Jones S, Rusche JR, Gottesfeld JM & Pook MA (2011) Prolonged treatment with pimelic o-aminobenzamide HDAC inhibitors ameliorates the disease phenotype of a Friedreich ataxia mouse model. *Neurobiol. Dis.* **42:** 496–505

Sandi C, Sandi M, Anjomani-Virmouni S, Al-Mahdawi S & Pook MA (2014a) Epigenetic-based therapies for Friedreich ataxia. *Front. Genet.* **5:** 165

Sandi C, Sandi M, Jassal H, Ezzatizadeh V, Anjomani-Virmouni S, Al-Mahdawi S & Pook MA (2014b) Generation and characterisation of Friedreich ataxia YG8R mouse fibroblast and neural stem cell models. *PLoS One* **9:** e89488

Sarsero J, Li L, Holloway T, Voullaire L, Gazeas S, Fowler K, Kirby D, Thorburn D, Galle A, Cheema S, Koenig M, Williamson R & Ioannou P (2004) Human BAC-mediated rescue of the Friedreich ataxia knockout mutation in transgenic mice. *Mamm. Genome* **15:** 370–382

Sarsero JP, Li L, Wardan H, Sitte K, Williamson R & Ioannou PA (2003) Upregulation of expression from the FRDA genomic locus for the therapy of Friedreich ataxia. *J. Gene Med.* **5:** 72–81

Saveliev A, Everett C, Sharpe T, Webster Z & Festenstein R (2003) DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing. *Nature* **422:** 909–913

Savouret C, Brisson E, Essers J, Kanaar R, Pastink A, te Riele H, Junien C & Gourdon G (2003) CTG repeat instability and size variation timing in DNA repair-deficient mice. *EMBO J.* **22:** 2264–2273

Scheuermann MO, Tajbakhsh J, Kurz A, Saracoglu K, Eils R & Lichter P (2004) Topology of genes and nontranscribed sequences in human interphase nuclei. *Exp. Cell Res.* **301:** 266–279

Schmucker S, Argentini M, Carelle-Calmels N, Martelli A & Puccio H (2008) The in vivo mitochondrial two-step maturation of human frataxin. *Hum. Mol. Genet.* **17:** 3521–3531

Schooley A, Vollmer B & Antonin W (2012) Building a nuclear envelope at the end of mitosis: coordinating membrane reorganization, nuclear pore complex assembly, and chromatin de-condensation. *Chromosoma* **121:** 539–554

Schulz JB, Boesch S, Burk K, Durr A, Giunti P, Mariotti C, Pousset F, Schols L, Vankan P & Pandolfo M (2009) Diagnosis and treatment of Friedreich ataxia: a European perspective. *Nat. Rev. Neurol.* **5:** 222–234

Shan Y, Napoli E & Cortopassi G (2007) Mitochondrial frataxin interacts with ISD11 of the NFS1/ISCU complex and multiple mitochondrial chaperones. *Hum. Mol. Genet.* **16:** 929–941

Sharma R, Bhatti S, Gomez M, Clark RM, Murray C, Ashizawa T & Bidichandani SI (2002) The GAA triplet-repeat sequence in Friedreich ataxia shows a high level of somatic instability in vivo, with a significant predilection for large contractions. *Hum. Mol. Genet.* **11:** 2175–2187

Sharma R, De Biase I, Gómez M, Delatycki MB, Ashizawa T & Bidichandani SI (2004) Friedreich ataxia in carriers of unstable borderline GAA triplet-repeat alleles. *Ann. Neurol.* **56:** 898–901

Shav-Tal Y, Darzacq X, Shenoy SM, Fusco D, Janicki SM, Spector DL & Singer RH (2004) Dynamics of single mRNPs in nuclei of living cells. *Science* **304:** 1797–1800

Sheinberger J & Shav-Tal Y (2013) The dynamic pathway of nuclear RNA in eukaryotes. *Nucleus* **4:** 195–205

Shishkin AA, Voineagu I, Matera R, Cherng N, Chernet BT, Krasilnikova MM, Narayanan V, Lobachev KS & Mirkin SM (2009) Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. *Mol. Cell* **35:** 82–92

Shopland LS, Johnson C V, Byron M, McNeil J & Lawrence JB (2003) Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains: evidence for local euchromatic neighborhoods. *J. Cell Biol.* **162:** 981–990

Soragni E, Chou CJ, Rusche JR & Gottesfeld JM (2015) Mechanism of action of 2-aminobenzamide HDAC inhibitors in reversing gene silencing in Friedreich's ataxia. *Front. Neurol.* **6:** 44

Soragni E, Herman D, Dent SYR, Gottesfeld JM, Wells RD & Napierala M (2008) Long intronic GAA•TTC repeats induce epigenetic changes and reporter gene silencing in a molecular model of Friedreich ataxia. *Nucleic Acids Res.* **36:** 6056–6065

Soragni E, Miao W, Iudicello M, Jacoby D, De Mercanti S, Clerico M, Longo F, Piga A, Ku S, Campau E, Du J, Penalver P, Rai M, Madara JC, Nazor K, O'Connor M, Maximov A, Loring JF, Pandolfo M, Durelli L, et al (2014) Epigenetic therapy for Friedreich ataxia. *Ann. Neurol.* **76:** 489–508

La Spada AR & Taylor JP (2010) Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nat. Rev. Genet.* **11:** 247–258

Spector DL (2006) SnapShot: cellular bodies. *Cell* **127:** 1071.e1–1071.e2

Spector DL & Lamond AI (2011) Nuclear speckles. *Cold Spring Harb. Perspect. Biol.* **3:** a000646

Sterner DE & Berger SL (2000) Acetylation of histones and transcription-related factors. *Microbiol. Mol. Biol. Rev.* **64:** 435–459

Strahl BD & Allis CD (2000) The language of covalent histone modifications. *Nature* **403:** 41–45

Sturm B, Bistrich U, Schranzhofer M, Sarsero JP, Rauen U, Scheiber-Mojdehkar B, de Groot H, Ioannou P & Petrat F (2005) Friedreich's ataxia, no changes in mitochondrial labile iron in human lymphoblasts and fibroblasts: a decrease in antioxidative capacity? *J. Biol. Chem.* **280:** 6701–6708

Takiyama Y, Igasrashi S, Rogaeva EA, Endo K, Rogaev EI, Tanaka H, Sherrington R, Sanpei K, Liang Y, Saito M, Tsuda T, Takano H, Ikeda M, Lin C, Chi H, Kennedy JL, Lang AE, Wherrett JR, Segawa M, Nomura Y, et al (1995) Evidence for inter-generational instability in the CAG repeat in the MJD1 gene and for conserved haplotypes at flanking markers amongst Japanese and Caucasian subjects with Machado-Joseph disease. *Hum. Mol. Genet.* **4:** 1137–1146

Takizawa T, Gudla PR, Guo L, Lockett S & Misteli T (2008a) Allele-specific nuclear positioning of the monoallelically expressed astrocyte marker GFAP. *Genes Dev.* **22:** 489–498

Takizawa T, Meaburn KJ & Misteli T (2008b) The meaning of gene positioning. *Cell* **135:** 9–13

Tambets K, Rootsi S, Kivisild T, Help H, Serk P, Loogväli E-L, Tolk H-V, Reidla M, Metspalu E, Pliss L, Balanovsky O, Pshenichnov A, Balanovska E, Gubina M, Zhadanov S, Osipova L, Damba L, Voevoda M, Kutuev I, Bermisheva M, et al (2004) The Western and Eastern roots of the Saami—the story of genetic 'outliers' told by mitochondrial DNA and Y chromosomes. *Am. J. Hum. Genet.* **74:** 661–682

Taroni F & DiDonato S (2004) Pathways to motor incoordination: the inherited ataxias. *Nat. Rev. Neurosci.* **5:** 641–655

Therizols P, Illingworth RS, Courilleau C, Boyle S, Wood AJ & Bickmore WA (2014) Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science* **346:** 1238–1242

Thomas EA, Coppola G, Desplats PA, Tang B, Soragni E, Burnett R, Gao F, Fitzgerald KM, Borok JF, Herman D, Geschwind DH & Gottesfeld JM (2008) The HDAC inhibitor 4b ameliorates the disease phenotype and transcriptional abnormalities in Huntington's disease transgenic mice. *Proc. Natl. Acad. Sci.* **105:** 15564–15569

Thomson I, Gilchrist S, Bickmore WA & Chubb JR (2004) The radial positioning of chromatin is not inherited through mitosis but is established de novo in early G1. *Curr. Biol.* **14:** 166–172

Thornton C, Wymer J, Simmons Z, McClain C & Moxley R (1997) Expansion of the myotonic dystrophy CTG repeat reduces expression of the flanking DMAHP gene. *Nat. Genet.* **16:** 407–409

Tomilin N V (1999) Control of genes by mammalian retroposons. *Int. Rev. Cytol.* **186:** 1–48

Tsukamoto T, Hashiguchi N, Janicki SM, Tumbar T, Belmont a S & Spector DL (2000) Visualization of gene activity in living cells. *Nat. Cell Biol.* **2:** 871–878

Vankan P (2013) Prevalence gradients of Friedreich's Ataxia and R1b haplotype in Europe co-localize, suggesting a common Palaeolithic origin in the Franco-Cantabrian ice age refuge. *J. Neurochem.* **126:** 11–20

Vargas DY, Raj A, Marras SAE, Kramer FR & Tyagi S (2005) Mechanism of mRNA transport in the nucleus. *Proc. Natl. Acad. Sci. U. S. A.* **102:** 17008–17013

Volpicelli-Daley LA, Lucast L, Gong L-W, Liu L, Sasaki J, Sasaki T, Abrams CS, Kanaho Y & De Camilli P (2010) Phosphatidylinositol-4-Phosphate 5-Kinases and Phosphatidylinositol 4,5-Bisphosphate synthesis in the brain. *J. Biol. Chem.* **285:** 28708–28714

Vyas PM, Tomamichel WJ, Pride PM, Babbey CM, Wang Q, Mercier J, Martin EM & Payne RM (2012) A TAT–Frataxin fusion protein increases lifespan and cardiac function in a conditional Friedreich's ataxia mouse model. *Hum. Mol. Genet.* **21:** 1230–1247

Walsh MJ, Cooper-Knock J, Dodd JE, Stopford MJ, Mihaylov SR, Kirby J, Shaw PJ & Hautbergue GM (2014) Decoding the pathophysiological mechanisms that underlie RNA dysregulation in neurodegenerative disorders: a review of the current state of the art. *Neuropathol. Appl. Neurobiol.* **41:** 109–134

Wang J, Shiels C, Sasieni P, Wu PJ, Islam SA, Freemont PS & Sheer D (2004) Promyelocytic leukemia nuclear bodies associate with transcriptionally active genomic regions. *J. Cell Biol.* **164:** 515–526

Weil TT, Parton RM & Davis I (2010) Making the message clear: visualizing mRNA localization. *Trends Cell Biol.* **20:** 380–390

Wells RD (2008) DNA triplexes and Friedreich ataxia. *FASEB J.* **22:** 1625–1634

Wilson KL & Foisner R (2010) Lamin-binding proteins. *Cold Spring Harb. Perspect. Biol.* **2:** a000554

Wu B, Chen J & Singer RH (2014) Background free imaging of single mRNAs in live cells using split fluorescent proteins. *Sci. Rep.* **4:** 3615

Xia H, Cao Y, Dai X, Marelja Z, Zhou D, Mo R, Al-Mahdawi S, Pook MA, Leimkühler S, Rouault TA & Li K (2012) Novel frataxin isoforms may contribute to the pathological mechanism of Friedreich ataxia. *PLoS One* **7:** e47847

Xu C, Soragni E, Chou CJ, Herman D, Plasterer HL, Rusche JR & Gottesfeld JM (2009) Chemical probes identify a role for histone deacetylase 3 in Friedreich's ataxia gene silencing. *Chem. Biol.* **16:** 980–989

Yamagata H, Miki T, Nakagawa M, Johnson K, Deka R & Ogihara T (1996) Association of CTG repeats and the 1-kbAlu insertion/deletion polymorphism at the myotonin protein kinase gene in the Japanese population suggests a common Eurasian origin of the myotonic dystrophy mutation. *Hum. Genet.* **97:** 145–147

Yanagisawa H, Fujii K, Nagafuchi S, Nakahori Y, Nakagome Y, Akane A, Nakamura M, Sano A, Komure O, Kondo I, Kyu Jin D, Sørensen SA, Potter NT, Robert Young S, Nakamura K, Nukina N, Nagao Y, Tadokoro K, Okuyama T, Miyashita T, et al (1996) A unique origin and multistep process for the generation of expanded DRPLA triplet repeats. *Hum. Mol. Genet.* **5:** 373–379

Yandim C, Natisvili T & Festenstein R (2013) Gene regulation and epigenetics in Friedreich's ataxia. *J. Neurochem.* **126:** 21–42

Ye Q, Callebaut I, Pezhman A, Courvalin J-C & Worman HJ (1997) Domain-specific interactions of human HP1-type chromodomain proteins and inner nuclear membrane protein LBR. *J. Biol. Chem.* **272:** 14983–14989

Yoder JA, Walsh CP & Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13:** 335–340

Yoon T & Cowan JA (2003) Iron−sulfur cluster biosynthesis. Characterization of frataxin as an iron donor for assembly of [2Fe-2S] clusters in ISU-type proteins. *J. Am. Chem. Soc.* **125:** 6078–6084

Yoon T & Cowan JA (2004) Frataxin-mediated iron delivery to ferrochelatase in the final step of heme biosynthesis. *J. Biol. Chem.* **279:** 25943–25946

Yoon T, Dizin E & Cowan JA (2007) N-terminal iron-mediated self-cleavage of human frataxin: regulation of iron binding and complex formation with target proteins. *JBIC J. Biol. Inorg. Chem.* **12:** 535–542

Yunger S, Rosenfeld L, Garini Y & Shav-Tal Y (2010) Single-allele analysis of transcription kinetics in living mammalian cells. *Nat. Methods* **7:** 631–633

Yunger S, Rosenfeld L, Garini Y & Shav-Tal Y (2013) Quantifying the transcriptional output of single alleles in single living mammalian cells. *Nat. Protoc.* **8:** 393–408

Zenklusen D, Larson DR & Singer RH (2008) Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.* **15:** 1263–1271

Zhang L-F, Huynh KD & Lee JT (2007) Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. *Cell* **129:** 693–706

Zhang Y, Muyrers JPP, Testa G & Stewart AF (2000) DNA cloning by homologous recombination in Escherichia coli. *Nat Biotech* **18:** 1314–1317

Zhao R, Bodnar MS & Spector DL (2009) Nuclear neighborhoods and gene expression. *Curr. Opin. Genet. Dev.* **19:** 172–179

Zimyanin VL, Belaya K, Pecreaux J, Gilchrist MJ, Clark A, Davis I & St Johnston D (2008) In vivo imaging of oskar mRNA transport reveals the mechanism of posterior localization. *Cell* **134:** 843–853

Zullo JM, Demarco IA, Piqué-Regi R, Gaffney DJ, Epstein CB, Spooner CJ, Luperchio TR, Bernstein BE, Pritchard JK, Reddy KL & Singh H (2012) DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina. *Cell* **149:** 1474–1487

ANNEXES

Publications

ANNEX 1

Lufino MMP, **Silva AM,** Nemeth AH, Alegre-Albarrategui J, Russell AJ, Wade-Martins R (2013) A GAA Repeat Expansion Reporter Model of Friedreich's Ataxia Recapitulates the Genomic Context and Allows Rapid Screening of Therapeutic Compounds. *Hum. Mol. Genet.* 22(25): 5173-87.

Human Molecular Genetics, 2013, Vol. 22, No. 25 **5173–5187**
doi:10.1093/hmg/ddt370
Advance Access published on August 13, 2013

A GAA repeat expansion reporter model of Friedreich's ataxia recapitulates the genomic context and allows rapid screening of therapeutic compounds

Michele M.P. Lufino[1], Ana M. Silva[1,2], Andrea H. Németh[3,4], Javier Alegre-Abarrategui[1,5], Angela J. Russell[6,7] and Richard Wade-Martins[1,5,*]

[1]Department of Physiology, Anatomy and Genetics, University of Oxford, Le Gros Clark Building, South Parks Road, Oxford OX1 3QX, UK, [2]Faculdade de Medicina, Universidade de Lisboa, Lisboa 1649-028, Portugal, [3]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford OX3 9DU, UK, [4]Department of Clinical Genetics, Churchill Hospital, Oxford University Hospitals NHS Trust, Oxford OX3 7LE, UK, [5]Oxford Parkinson's Disease Centre, University of Oxford, Le Gros Clark Building, South Parks Road, Oxford OX1 3QX, UK, [6]Department of Chemistry, Chemistry Research Laboratory and [7]Department of Pharmacology, University of Oxford, Mansfield Road, Oxford OX1 3QT, UK

Received May 20, 2013; Revised July 15, 2013; Accepted July 26, 2013

Friedreich's ataxia (FRDA) is caused by large GAA expansions in intron 1 of the frataxin gene (*FXN*), which lead to reduced *FXN* expression through a mechanism not fully understood. Understanding such mechanism is essential for the identification of novel therapies for FRDA and this can be accelerated by the development of cell models which recapitulate the genomic context of the *FXN* locus and allow direct comparison of normal and expanded *FXN* loci with rapid detection of frataxin levels. Here we describe the development of the first GAA-expanded *FXN* genomic DNA reporter model of FRDA. We modified BAC vectors carrying the whole *FXN* genomic DNA locus by inserting the luciferase gene in exon 5a of the *FXN* gene (pBAC-*FXN-Luc*) and replacing the six GAA repeats present in the vector with an ~310 GAA repeat expansion (pBAC-*FXN-GAA-Luc*). We generated human clonal cell lines carrying the two vectors using site-specific integration to allow direct comparison of normal and expanded *FXN* loci. We demonstrate that the presence of expanded GAA repeats recapitulates the epigenetic modifications and repression of gene expression seen in FRDA. We applied the GAA-expanded reporter model to the screening of a library of novel small molecules and identified one molecule which up-regulates *FXN* expression in FRDA patient primary cells and restores normal histone acetylation around the GAA repeats. These results suggest the potential use of genomic reporter cell models for the study of FRDA and the identification of novel therapies, combining physiologically relevant expression with the advantages of quantitative reporter gene expression.

INTRODUCTION

Friedreich's ataxia (FRDA; OMIM 229300) is a progressive neurodegenerative disorder and the most common form of recessive ataxia, affecting approximately 1–2 in 50 000 Caucasians (1). Patients present with progressive gait and limb ataxia, lower limb areflexia, dysarthria, increased incidence of diabetes and hypertrophic cardiomyopathy, which subsequently leads to death in the fourth or fifth decade of life (2,3). The neurological symptoms are mainly caused by degeneration of the large sensory neurons of the dorsal root ganglia, the spinocerebellar tracts and the dentate nucleus of the cerebellum (4,5). FRDA is caused by an abnormal expansion of GAA repeats in intron 1 of the frataxin gene (*FXN*) (1). Approximately 98% of FRDA patients are homozygous for a GAA repeat expansion and the remaining patients are compound heterozygous with one

*To whom correspondence should be addressed. Tel: +44 1865282837; Fax: +44 1865272420; Email: richard.wade-martins@dpag.ox.ac.uk

expanded allele and a point mutation in the second allele (6,7). Normal unaffected individuals have <36 GAA repeats, whereas FRDA patients present with GAA expansions ranging from 70 to >1000 GAA repeats which lead to reduced levels of frataxin, a nuclear-encoded mitochondrial protein essential for life (1,8). The GAA size of the small allele has been shown to correlate with residual frataxin levels, earlier onset and increased severity of disease (9,10). Frataxin deficiency leads to iron–sulphur cluster deficiency, mitochondrial iron accumulation and increased susceptibility to oxidative stress (11–16).

The mechanism through which expanded GAA repeats silence *FXN* expression still needs further elucidation. Two non-exclusive models have been proposed (11,17). Initial evidence suggested that expanded GAA repeats in intron 1 of *FXN* form unusual DNA structures such as triplexes or sticky DNA and DNA/RNA hybrid structures, which impede the progress of the RNA polymerase and perturb transcription in a length-dependent manner (18–24). However, more recently, a second model suggests that long GAA expansions can induce silencing of *FXN* expression via a heterochromatin-mediated mechanism of repression (25,26). Epigenetic changes around expanded GAA repeats have been identified, which include increased DNA methylation at specific CpG sites upstream of the GAA repeats (27–30) and reduced acetylation of histones H3 and H4 accompanied by increased levels of methylated histones H3K9me2 and H3K9me3 in regions flanking GAA repeats (26,31). The *FXN* promoter in patient-derived cells and tissues shows a less permissive configuration for transcription initiation (27,32). More recently, a depletion of chromatin insulator protein CTCF was identified at the *FXN* promoter of FRDA patient-derived cells and a correlation between CTCF depletion and increased levels of the frataxin antisense transcript-1 was suggested (33).

Currently, there is no proven treatment for FRDA, although there are promising therapies under development (26,34–37). A better understanding of the *FXN* silencing which occurs in the presence of large GAA expansions is vital for the identification of novel therapies for FRDA. The development of reporter models which reproduce the epigenetic hallmarks of FRDA while providing efficient ways to quantify *FXN* expression would considerably accelerate the identification of such treatments. A few GAA-based reporter models have been described; however, these focus only on the use of short heterologous reporter constructs carrying expanded GAA repeats out of context and lacking *FXN* genomic DNA sequences (31,38,39). Such models do not carry repeat expansions within the *FXN* locus and thus they do not allow the analysis of the nature of the *FXN* silencing induced by long GAA repeats. A reporter model based on the use of the whole *FXN* genomic DNA locus would provide instead an excellent tool for such study since the expansion would be present within its natural genomic context, within intron 1 of the *FXN* gene. Furthermore, such reporter models achieve physiologically relevant *FXN* expression, since the native promoter and all the regulatory elements necessary for physiological gene expression are present in the vector (40–42).

Here we describe the development and characterization of the first GAA-expanded genomic DNA reporter model of FRDA. Using homologous recombination, we modified a BAC carrying the 80 kb *FXN* locus by inserting the reporter gene luciferase in exon 5a of the *FXN* gene, generating the pBAC-*FXN-Luc* vector. We also replaced a normal number of GAA repeats (six GAA) present in pBAC-*FXN-Luc* with ~310 GAA repeats (pBAC-*FXN-GAA-Luc* vector) and generated stable clonal cell lines in order to compare the effect of normal or expanded GAA repeats on the *FXN* gene. Using site-specific integration coupled with a copy number assay, we generated highly comparable cell lines which allow comparison in the absence of confounding effects. We show that ~310 GAA repeats decrease *FXN* expression by greatly affecting the epigenetic landscape of the *FXN* locus. Finally, we applied this novel reporter model to the screening of a library of small molecules and identified a molecule which up-regulates *FXN* expression in FRDA patient cells and reverses the GAA-induced *FXN* gene silencing by increasing histone acetylation at the *FXN* locus. We believe that our genomic DNA reporter model allows accurate comparison of normal and GAA-expanded *FXN* loci within a physiological genomic context and provides a tool for rapid and cost-effective identification of *FXN* up-regulating therapeutic strategies.

RESULTS

Construction of pBAC-*FXN-Luc* and pBAC-*FXN-GAA-Luc* genomic DNA reporter vectors

A BAC clone (RP11-265B8) containing the whole *FXN* locus from human chromosome 9 was used to generate *FXN-Luc* BAC vectors. This vector contains approximately 38 kb of promoter region, the 80 kb *FXN* locus and 17 kb of downstream sequence (42). Since the most abundant *FXN* transcript consists of exons 1–5a (1,43), we inserted the firefly luciferase sequence in exon 5a immediately prior to the stop codon, generating pBAC-*FXN-Luc* fusion vector (Fig. 1A). We included a sequence encoding a five amino-acid Gly-Ser-Gly-Ser-Gly (GSGSG) peptide linker between the exon 5a and luciferase sequence (Fig. 1A) to allow correct folding of the frataxin and luciferase proteins; the choice of the composition of the linker was based on previous reports (44). Construction of pBAC-*FXN-Luc* was achieved by a selection/counter-selection homologous recombination strategy using homology arms to exon 5a. To confirm correct vector recombination, successful colonies were analysed by junction PCR, by restriction enzyme digestion followed by pulsed-field gel electrophoresis (PFGE) and by sequencing, which confirmed correct insertion of the GSGSG-luciferase sequence in exon 5a (data not shown). The resulting modified *FXN* gene expresses an FXN-luciferase fusion protein within the context of the *FXN* genomic DNA locus.

In order to insert a GAA expansion into intron 1 of the pBAC-*FXN-Luc* vector, a PCR product containing expanded GAA repeats was amplified from genomic DNA isolated from an FRDA patient-derived lymphoblastoid cell line (GM16207, alleles with 280 and 830 GAA repeats in the *FXN* gene) using an established PCR protocol (1,45). The smaller PCR product carrying ~280 GAA repeats flanked by 195 and 255 bp of homology arms upstream and downstream of the GAA repeats, respectively, was preferentially amplified and used for recombination. We used the RPSL-Neo selection/counter-selection homologous recombination method to replace the six GAA repeats present in the *FXN*-BAC in intron 1 of *FXN-Luc* gene with the expanded GAA repeats amplified by PCR, generating pBAC-*FXN-GAA-Luc* vector. Owing to low efficiency of recombination, successful colonies were identified through

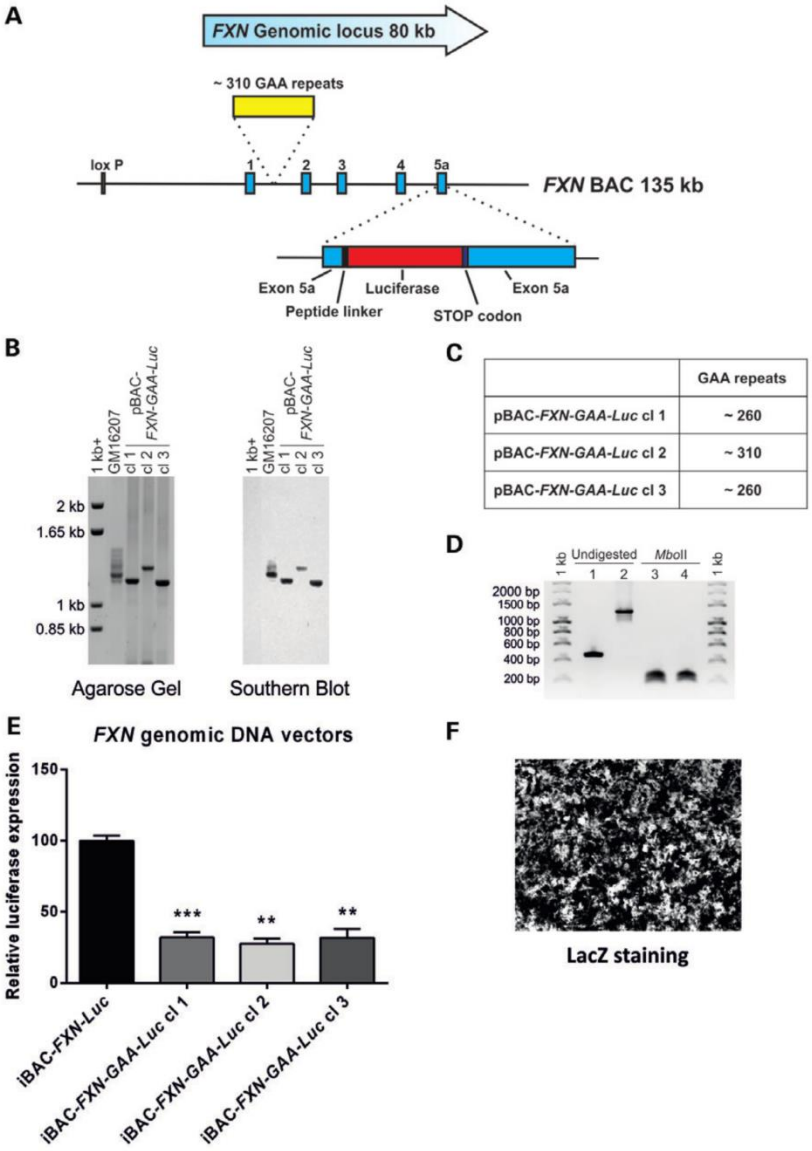

Figure 1. Generation of pBAC-*FXN-Luc* and pBAC-*FXN-GAA-Luc* genomic DNA reporter vectors. (**A**) Schematic representation of the construction of pBAC-*FXN-Luc* and pBAC-*FXN-GAA-Luc* vectors. Construction was achieved in two successive rounds of selection–counter-selection homologous recombination. First, a luciferase sequence preceded by a GSGSG peptide linker was introduced at the 5′ end of exon 5a, immediately before the stop codon, generating pBAC-*FXN-Luc* vector. This vector expresses an FXN-luciferase fusion protein and carries six GAA repeats in intron 1. To generate pBAC-*FXN-GAA-Luc* vector, a second recombination was performed to replace the six GAA repeats present in intron 1 with ∼310 GAA repeats amplified from FRDA patient-derived cells. (**B**) Successful insertion of expanded GAA repeats was confirmed by Southern blot by using a DIG-labelled TTC_{10} probe. Comparison with ladder allowed sizing of GAA repeats as reported in (**C**). (C) Size of GAA expansions in pBAC-*FXN-GAA-Luc* clones. (**D**) pBAC-*FXN-GAA-Luc* cl 2 vector carries ∼310 pure GAA repeats. GAA repeats were amplified from pBAC-*FXN-Luc* (lane 1) and pBAC-*FXN-GAA-Luc* (lane 2) vectors, incubated with *Mbo*II and run on a 1% agarose gel. The PCR products of pBAC-*FXN-Luc* (lane 3) and pBAC-*FXN-GAA-Luc* (lane 4) are fully digested by the enzyme, leaving the flanking regions of 208 and 248 bp and demonstrating the lack of interruptions in the GAA expansion. (**E**) To assess whether expanded GAA repeats reduce FXN-luciferase expression, we delivered iBAC-*FXN-Luc* and iBAC-*FXN-GAA-Luc* vectors to the neuronal cell line SH-SY5Y using HSV-1 amplicon vectors. Luciferase assay showed a 70–75% reduction in FXN-luciferase levels in the three iBAC-*FXN-GAA-Luc* clones. Error bars represent mean ± SEM ($n = 3$). $^{**}P < 0.01$, $^{***}P < 0.001$ as determined by the one-way ANOVA compared with iBAC-*FXN-Luc*, with Dunnett's test. (**F**) LacZ staining of the experiment described in (E) shows high efficiency of vector delivery to SH-SY5Y cells.

colony blot using a TTC_{10} probe. The GAA expansion of three independent colonies was sized through Southern blotting, which confirmed an insertion of up to ~310 GAA repeats (Fig. 1B and C). To assess the purity of the ~310 GAA repeats, sequencing was performed from both sides. We confirmed the presence of 213 GAA repeats on the GAA strand and 125 GAA repeats on the TTC strand, thereby achieving an overlap of at least 28 GAA repeats (data not shown). The presence of such overlap ensured that the whole GAA expansion was sequenced. We detected no interruptions in the expanded GAA sequence and this was further corroborated by *Mbo*II digestion according to a method previously described (Fig. 1D) (46). The GAA repeats were amplified from pBAC-*FXN-GAA-Luc* cl 2 vector by PCR and digested with *Mbo*II. The complete digestion of the PCR product confirms the purity of the GAA expansion in the pBAC-*FXN-GAA-Luc* cl 2 vector (Fig. 1D).

Validation of pBAC-*FXN-Luc* and pBAC-*FXN-GAA-Luc* reporter constructs

To investigate the effect of ~310 GAA repeats on FXN-luciferase expression in the human neuronal cell line SH-SY5Y, we used the herpes simplex virus type-1 (HSV-1) amplicon vector system. SH-SY5Y cells are characterized by low efficiency of transfection and HSV-1 vectors mediate intact delivery of BACs to cells at high efficiency (40,47). To allow packaging into HSV-1 amplicons, we used *Cre/loxP* recombination (Fig. 2A) to incorporate into pBAC-*FXN-Luc* and pBAC-*FXN-GAA-Luc* a retrofitting vector containing the HSV-1 *oriS* origin of replication and *pac* packaging signals, a *lacZ* cassette for titration and sequences derived from the Epstein–Barr virus for extra-chromosomal vector retention, generating iBAC-*FXN-Luc* and iBAC-*FXN-GAA-Luc*, respectively. Correctly retrofitted vectors were identified using PFGE analysis, and vectors were packaged into HSV-1 amplicons using an improved helper virus-free packaging protocol as described previously (data not shown) (48). Average titres of $1-2 \times 10^7$ transducing units per millilitre were obtained for all vectors. SH-SY5Y cells were transduced with iBAC-*FXN-Luc* and iBAC-*FXN-GAA-Luc* cl 1, 2 and 3 amplicons at a multiplicity of infection of 4 and luciferase assay was performed 4 days after infection. High efficiency of transduction was achieved as determined by LacZ staining (Fig. 1F). The presence of up to ~310 GAA repeats in iBAC-*FXN-GAA-Luc* vectors causes a reduction in FXN-luciferase expression by ~75% when compared with iBAC-*FXN-Luc* vector, recapitulating the effect of GAA repeats on *FXN* expression observed in FRDA patient cells (Fig. 1E). pBAC-*FXN-GAA-Luc* cl 2 vector (referred to as pBAC-*FXN-GAA-Luc*) was chosen for the following experiments.

Generation and characterization of a GAA-expanded genomic DNA reporter model of FRDA

In order to generate a cell model which allows the dissection of the effect of GAA repeats on *FXN* expression, we generated stable clonal cell lines carrying pBAC-*FXN-Luc* and pBAC-*FXN-GAA-Luc* vectors by using site-specific vector integration, since random vector integration in the genome can affect transgene expression levels. Site-specific integration allows precise comparison of the two vectors in the absence of confounding effects due to differential integration site, since the two vectors are integrated at the same genomic location. To generate these cell lines, we adapted the previously described Flp-In system (Life Technologies) to BAC vectors. We developed an Flp-In BAC integration system by generating the retrofitting vector pH-FRT-Hy (Fig. 2A), which contains the Flp-In promoter-less hygromycin cassette. pH-FRT-Hy was then retrofitted into the pBAC-*FXN-Luc* and pBAC-*FXN-GAA-Luc* vectors using the *Cre/loxP* retrofitting strategy previously described (40), generating pFRT-*FXN-Luc* and pFRT-*FXN-GAA-Luc* vectors (Fig. 2A). A stable FRT acceptor cell line was generated by transfecting HEK cells with the plasmid pFRT-LacZeo, followed by zeocin selection and confirmation of positive LacZ staining (HEK FRT cells). These cells were transfected with pFRT-*FXN-Luc* and pFRT-*FXN-GAA-Luc* vectors together with the Flp recombinase-encoding plasmid pOG44 (Life Technologies). Stable clonal cell lines (referred to as *FXN-Luc* and *FXN-GAA-Luc*) were isolated in the presence of hygromycin, and PCR was used to confirm correct vector integration at the FRT acceptor site (Fig. 2B). Random integration events are prevented since the promoter which drives hygromycin is only present at the docking site and any integration at other genomic locations will not result in antibiotic resistance. The presence of GAA repeats in *FXN-GAA-Luc* cells was confirmed by PCR (data not shown). Clonal cell lines were expanded and characterized.

Fluorescence *in situ* hybridization (FISH) was performed on *FXN-Luc* and *FXN-GAA-Luc* clonal cell lines using the unmodified *FXN* BAC and the pH-FRT-Hy retrofitting vector as probes which, when co-localizing, indicate the presence of the integrated vector (Fig. 2C). *FXN-Luc* and *FXN-GAA-Luc* cell lines, which have been generated using different vectors, both show vector integration at the same location, on chromosome 1p as confirmed by the use of a chromosome 1 centromeric probe (data not shown), confirming the consistency of the site-specific integration for large genomic DNA vectors (Fig. 2C). We then assessed vector copy number in *FXN-Luc* and *FXN-GAA-Luc* cells by real-time PCR, using as reference sample the acceptor cell line HEK FRT, which carries three endogenous *FXN* loci as determined by FISH (Fig. 2C). We used this assay to select *FXN-Luc* and *FXN-GAA-Luc* clonal cell lines carrying one copy of transgene (Fig. 2D).

We then determined the effect of the GAA repeat expansion on FXN-luciferase expression by reverse transcription real-time PCR (qRT-PCR) and by luciferase assay. The presence of GAA repeats causes a reduction of 37% in FXN-luciferase mRNA levels (Fig. 2F) and 42% in FXN-luciferase protein levels (Fig. 2G), as determined by qRT-PCR and by luciferase assay, respectively. Western blot analysis shows that *FXN-Luc* and *FXN-GAA-Luc* cells express an FXN-luciferase fusion protein of the expected size of ~79 kDa (Fig. 2E).

Epigenetic characterization of *FXN-Luc* and *FXN-GAA-Luc* cell lines

Recently, it has been shown that the presence of GAA repeats in intron 1 of the *FXN* locus reduces *FXN* expression by heterochromatin formation and increased CpG methylation around GAA

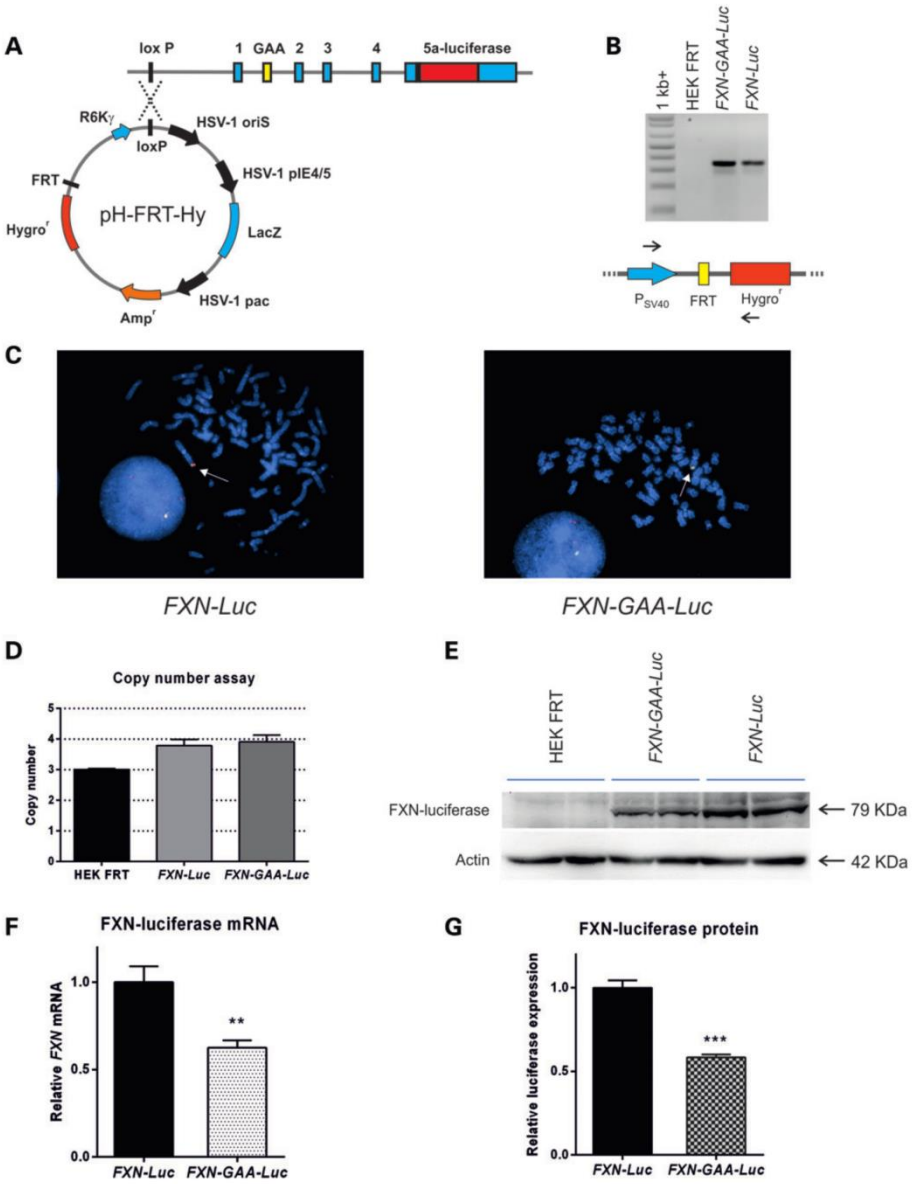

Figure 2. Generation and characterization of a GAA-expanded genomic DNA reporter model of FRDA. (**A**) Schematic representation of *Cre/loxP* retrofitting of pBAC-*FXN-Luc* and pBAC-*FXN-GAA-Luc* vectors with pH-FRT-Hy. pH-FRT-Hy carries a promoter-less hygromycin cassette preceded by an FRT site. (**B**) Site-specific vector integration in *FXN-Luc* and *FXN-GAA-Luc* cells is confirmed by PCR using primers targeting the P_{SV40} promoter and the hygromycin cassette. (**C**) FISH analysis was performed on a representative clonal cell line for each vector, using the unmodified *FXN*-BAC (red) and the retrofitting vector pH-FRT-Hy (green) as probes. Co-localization of the two probes (arrows) shows vector integration in chromosome 1p. Chromosome identity was confirmed using a chromosome 1 centromeric probe (data not shown). The cells analysed here carry three copies of the endogenous *FXN* locus, due to the hypotriploid nature of HEK cells. (**D**) Copy number was used to identify stable cell lines with one vector copy. Copy number was determined by real-time PCR using UpGAA primers and data normalized by *GAPDH*. Data are expressed as relative to HEK FRT, which was set to three copies based on the identification in (C) of three endogenous *FXN* loci. (**E**) Western blot analysis of *FXN-Luc* and *FXN-GAA-Luc* clonal cell lines using an anti-luciferase antibody shows a ~79 kDa band which corresponds to the expected size for the FXN-luciferase fusion protein. (**F**) qRT-PCR in *FXN-Luc* and *FXN-GAA-Luc* clonal cell lines using exon 5a-Luc primers shows a 37% reduction in FXN-luciferase mRNA levels in *FXN-GAA-Luc* cells. Exon 5a-Luc data are normalized to *GAPDH* and expressed as relative to *FXN-Luc* cells. Error bars represent mean ± SEM ($n = 4$). $^{**}P < 0.01$, as determined by Student's *t*-test. (**G**) Luciferase assay shows a 42% reduction in FXN-luciferase protein levels. Data are expressed as relative light units and normalized to total cell protein. Error bars represent mean ± SEM ($n = 3$). $^{***}P < 0.001$, as determined by Student's *t*-test.

repeats (26,27). To test for the presence of repressive epigenetic hallmarks in our cell model, we first analyzed histone modifications by chromatin immunoprecipitation (ChIP) at three sites, at the promoter and the regions flanking GAA repeats, upstream and downstream of GAA repeats, using antibodies specific for the human acetylated histones H3K9 and H4K8 and di- and tri-methylated histone H3K9. This analysis revealed a decrease in histone acetylation and an increase in histone methylation across the three regions in the *FXN-GAA-Luc* cell line when compared with *FXN-Luc* cells (Fig. 3A), as previously reported in FRDA patient brain tissue and patient-derived cell lines (26,27,32).

We then analyzed CpG methylation by bisulfite sequencing as reported in previous studies (27). In the region upstream of the expanded GAA repeats, we found increased DNA methylation in *FXN-GAA-Luc* cells at CpG sites 4 and 5 (Fig. 3B), in agreement with previously published methylation data from FRDA patients (27,29). Our results for CpG 6 differ from the results previously found in patients (27,28), but are in agreement with methylation data from the brain of FRDA mice (27). In the region downstream of the GAA repeats, we observed increased DNA methylation at CpG sites 1, 2 and 7 (Fig. 3B).

Since *FXN-Luc* and *FXN-GAA-Luc* cell lines share the same integration site and only differ by the presence of an ~310 GAA repeats expansion, it is most likely that the GAA expansion is causing the changes in histone acetylation/methylation and DNA methylation observed (Fig. 3).

Use of *FXN-GAA-Luc* cell line for the screening of chemical libraries

We have shown above that *FXN-GAA-Luc* cells carry a GAA expansion which causes heterochromatin-mediated silencing of FXN-luciferase expression. Since recent publications report

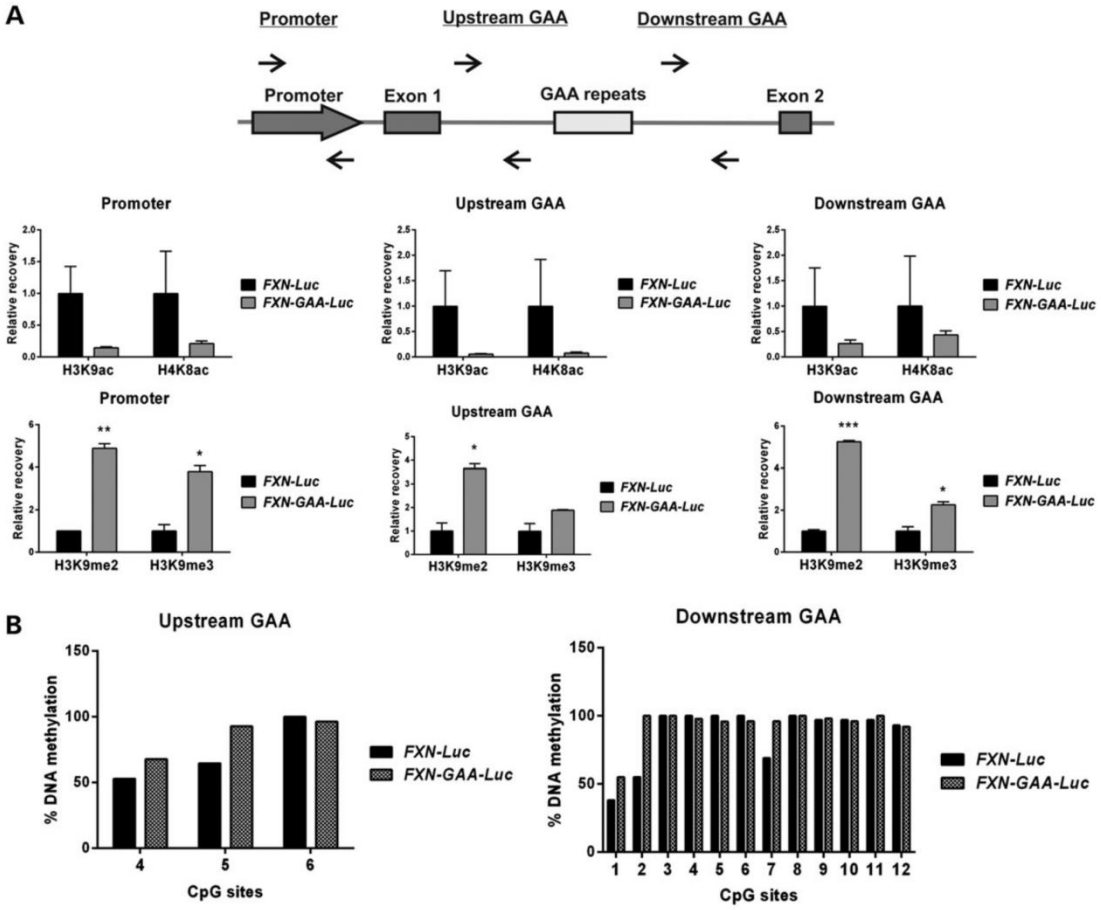

Figure 3. Expanded GAA repeats induce heterochromatin-mediated *FXN* silencing. (**A**) ChIP was performed on *FXN-Luc* and *FXN-GAA-Luc* cells using antibodies specific for the human acetylated histones H3K9 and H4K8 and di- and tri-methylated histone H3K9. The schematic diagram shows the position of the primers used. Error bars represent mean ± SEM from two independent immunoprecipitations and each immunoprecipitation quantified in triplicate. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ as determined by Student's *t*-test. (**B**) CpG methylation was analysed upstream and downstream of GAA repeats using bisulfite sequencing. We found increased DNA methylation at CpG sites 4 and 5 upstream of GAA repeats and at CpG sites 1, 2 and 7 downstream of GAA repeats. A total of 10 colonies were sequenced per region and for each cell line.

the successful up-regulation of *FXN* expression with histone deacetylase (HDAC) inhibitors (26,34), we applied our novel genomic DNA reporter model of FRDA to the screening of novel small molecules with potential HDAC inhibitor function. We screened *in silico* a library of 25 000 compounds to extract representative structures incorporating known pharmacophores associated with HDAC inhibitors and, more generally, with Zn(II)-binding motifs. This library has been designed to cover a wide range of biological space, including pharmacophores with well-characterized biological mechanisms in addition to structural motifs which exhibit biological effects with an unknown mechanism (49,50). Structures have been excluded which contain highly reactive functionalities (e.g. aldehydes, or Schiff bases) or known toxicophores (e.g. poly-halogenated species or poly-nitro aromatics) and selected structures are amenable to both resynthesis and rapid diversification. Compounds were selected from this library based on the presence of motifs likely to bind to Zn(II), such as hydroxamic acids, diamines and amino alcohols, which may plausibly inhibit zinc-dependent enzymes such as HDACs. However, alternative mechanisms of action cannot be ruled out. We identified 88 potential Zn(II)-binding compounds and performed the screening of such molecules in a 96-well format in triplicate by incubating *FXN-GAA-Luc* cells at a standard concentration of 20 μM for 48 h and assessing FXN-luciferase expression by luciferase assay (Fig. 4A). We identified four compounds which significantly increased *FXN-GAA-Luc* expression levels above dimethyl sulphoxide (DMSO) levels. In order to exclude the possibility of interaction of compounds with the luciferase assay and to discard those that show a generalized unspecific increase of gene expression, we tested the effect of the selected four compounds on luciferase expression driven by the ubiquitous cytomegalovirus (CMV) promoter (Fig. 4B). None of the four compounds shows an increase in luciferase expression; however, three compounds show a decrease in luciferase levels, which could indicate early signs of cell toxicity. To avoid potentially cytotoxic molecules, we focused subsequent studies on compound C5, the only molecule which did not affect CMV-luciferase levels (Fig. 4B).

To assess whether C5 acts as an HDAC inhibitor on the *FXN* gene and is able to reverse the GAA-mediated *FXN* silencing, we incubated FRDA patient lymphoblastoid cells with C5 at 20 μM for 48 h and analysed H3K9 and H4K8 acetylation upstream and downstream of GAA repeats. The FRDA cell line GM15850 shows a reduction in histone acetylation at these areas when compared with the wild-type cell line GM15851. When FRDA patient cells are incubated with C5, the FRDA histone acetylation is restored to wild-type levels, suggesting C5 acts either directly or indirectly in inhibiting HDAC activity (Fig. 4C). C5 is an amino alcohol and its structure is shown in Figure 4D. To confirm the activity of C5, we sourced an authentic sample and further corroborated its identity through mass spectrometry.

Characterization of C5 on *FXN* expression

In order to further characterize this novel compound, we then performed a dose–response assay by incubating *FXN-GAA-Luc* cells

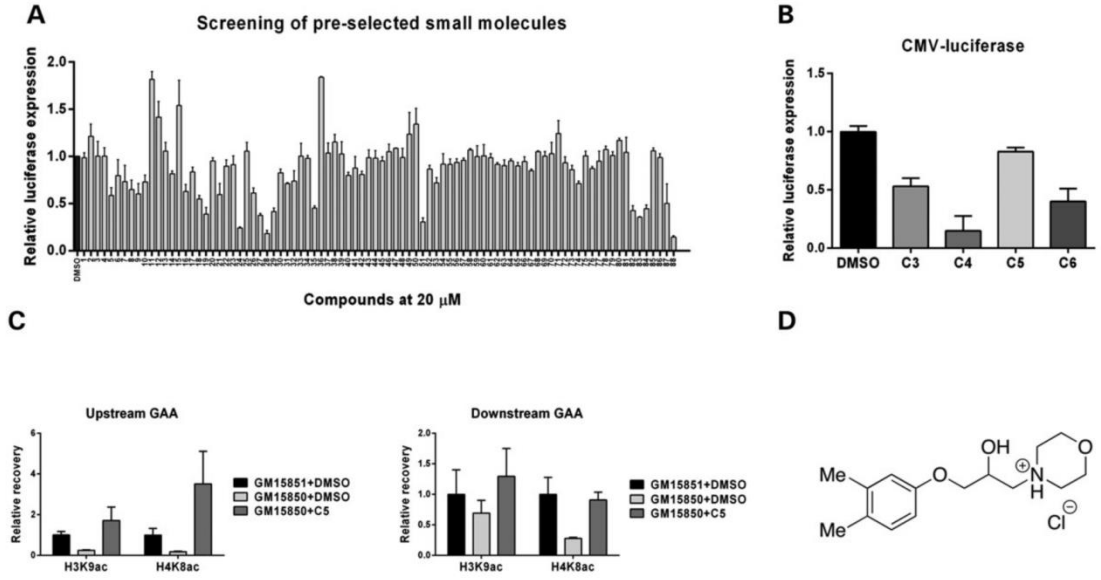

Figure 4. Screening of 88 pre-selected small molecules. (**A**) *FXN-GAA-Luc* cells were incubated in a 96-well format for 48 h with each compound at a final concentration of 20 μM. FXN-luciferase levels were quantified through luciferase assay. Each compound was tested in triplicate. (**B**) Four compounds identified from the screen were tested on HEK cells transfected with a CMV-luciferase plasmid to exclude molecules which directly affect the reporter assay or cause unspecific increase of expression. Transfected cells were incubated with each compound at a concentration of 20 μM for 48 h, followed by luciferase assay. (**C**) C5 increases H3K9 and H4K8 histone acetylation in FRDA lymphoblastoid cells (GM15850) to the normal levels observed in control cells (GM15851). Cells were incubated with C5 at 20 μM for 48 h and ChIP was performed using antibodies specific for the H3K9ac and H4K8ac residues. Error bars represent mean ± SEM from three independent immunoprecipitations and each immunoprecipitation quantified in triplicate. (**D**) Structure of the amino alcohol C5.

with C5 at concentrations ranging from 1 μM to 1 mM for 48 h. A sigmoidal-shaped dose-dependent increase in *FXN-GAA-Luc* expression was observed, with a high up-regulating effect observed at 300, 600 and 1000 μM (Fig. 5A). These three doses were chosen for further studies.

In order to determine whether the increase in *FXN-GAA-Luc* expression is due to an increase in transcription, we incubated *FXN-GAA-Luc* cells with C5 at the three chosen concentrations and analysed *FXN-GAA-Luc* mRNA by qRT-PCR. We show that C5 significantly increases *FXN-GAA-Luc* mRNA levels at 600 and 1000 μM (Fig. 5B). We then compared the effect of C5 on normal and GAA-expanded *FXN* loci, by incubation of *FXN-Luc* and *FXN-GAA-Luc* cells with C5 for 48 h followed by luciferase assay. We found that C5 induces a significantly greater up-regulating effect on the *FXN-GAA-Luc* cells, demonstrating a relative specificity for the expanded *FXN* locus (Fig. 5C). However, a small but significant increase was also observed on *FXN-Luc* cells.

Finally, to exclude a specific action of C5 on the transgene only, we incubated untransfected HEK cells with C5 at the three concentrations for 48 h and quantified endogenous *FXN* mRNA levels. We observed a 1.5–2-fold increase in endogenous *FXN* mRNA at 600 and 1000 μM (Fig. 5D). Furthermore, C5 significantly increases endogenous frataxin protein levels by 1.2-fold at 600 μM and by 1.7-fold at 1000 μM, as determined by frataxin dipstick assay (Fig. 5E).

Validation of C5 in primary cells from FRDA patients

To test the efficacy of C5 as a potential novel therapeutic molecule for FRDA, we tested C5 on primary lymphocytes isolated from FRDA patients, as these cells provide a readily obtainable source of patient primary cells. Primary lymphocytes were isolated from blood samples from three patients, using the Ficoll-Paque gradient as previously described (51). We incubated isolated primary lymphocytes with C5 at concentrations ranging from 300 to 800 μM and after 72 h analyzed *FXN* mRNA levels by qRT-PCR. In order to reduce the influence of fluctuations of reference genes on the final data, we normalized the *FXN* mRNA levels by the geometric mean of three different reference genes. Incubation with C5 up-regulates *FXN* expression in all three patients by 1.5–2-fold (Fig. 6), an increase similar to that observed in *FXN-GAA-Luc* cells. The identification of a new molecule through our cell model and the confirmation of its effect in primary lymphocytes from FRDA patients highlight the suitability of the *FXN-GAA-Luc* cells as a high-throughput expression model of FRDA.

DISCUSSION

Here we describe the generation of a novel genomic DNA reporter model of FRDA which allows direct comparison between normal and GAA-expanded genomic DNA loci and provides

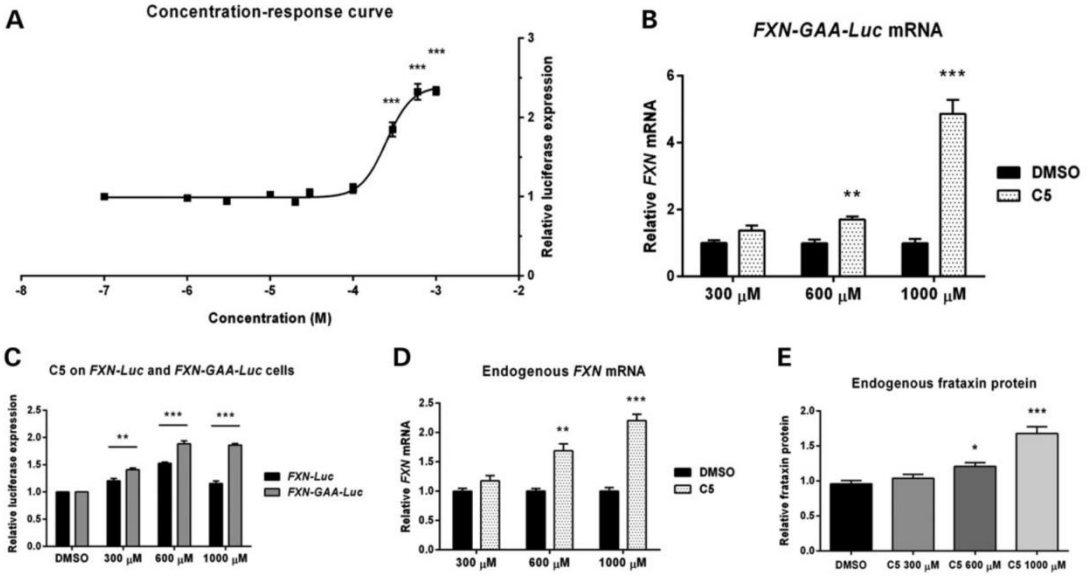

Figure 5. Effect of C5 on *FXN* expression in HEK cells. (**A**) Dose response of C5 on *FXN-GAA-Luc* cells. *FXN-GAA-Luc* cells were incubated for 48 h and luciferase assay was carried out. We identified 300, 600 and 1000 μM as the doses achieving the highest increase in FXN-luciferase expression. Error bars represent mean ± SEM ($n = 3$). $^{***}P < 0.001$ as determined by the one-way ANOVA compared with DMSO, with Dunnett's test. (**B**) qRT-PCR on *FXN-GAA-Luc* cells incubated with C5 at the three concentrations for 48 h shows a significant increase in *FXN-GAA-Luc* mRNA levels at 600 and 1000 μM. Data are normalized to *GAPDH* mRNA levels. (**C**) Luciferase assay on *FXN-Luc* and *FXN-GAA-Luc* cells after 48 h incubation with C5 at 300, 600 and 1000 μM. C5 shows a greater up-regulation on *FXN-GAA-Luc* cells at all three concentrations, although a smaller but significant increase in *FXN-Luc* cells is also observed. Data for each cell line are normalized by their respective DMSO. (**D**) qRT-PCR on HEK cells incubated with C5 for 48 h shows that C5 is active on the *FXN* endogenous locus. (**E**) C5 significantly increases endogenous frataxin protein levels by 1.2-fold at 600 μM and by 1.7-fold at 1000 μM as determined by frataxin dipstick assay. In (B)–(E), error bars represent mean ± SEM. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ as determined by Student's *t*-test (B)–(D) and by the one-way ANOVA compared with DMSO, with Dunnett's test (E).

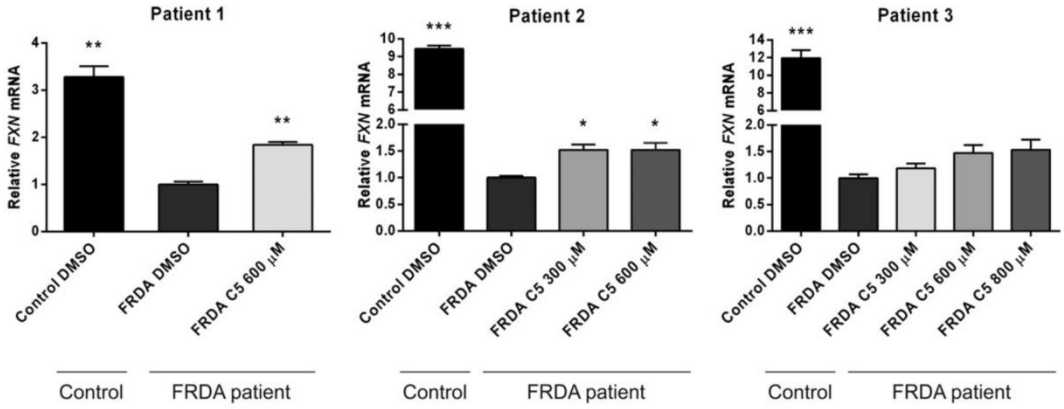

Figure 6. Validation of C5 on primary lymphocytes from FRDA patients. FRDA patient primary lymphocytes were extracted from blood using a Ficoll-Paque gradient and incubated with C5 for 72 h at different concentrations. C5 significantly increases *FXN* expression in patient 1 and 2 and shows a dose-dependent trend in patient 3. *FXN* data are normalized by the geometric mean of three reference genes: *GAPDH*, *HPRT* and *Beta-Actin*. Error bars represent mean ± SEM ($n = 4$). $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ as determined by the one-way ANOVA compared with FRDA DMSO, with Dunnett's test.

rapid and precise quantitation of *FXN* expression. To generate this reporter model, we used BAC vectors carrying the whole 80 kb genomic DNA *FXN* locus, since BAC vectors provide a suitable tool to study the molecular behaviour of expanded GAA repeats within their natural genomic context.

We have previously shown the importance of driving transgene expression from the native *FXN* promoter instead of from strong heterologous promoters (52) and we have demonstrated that by delivering a BAC clone (RP11-265B8) containing the whole 80 kb *FXN* locus to FRDA patient primary fibroblasts, we could rescue the increased sensitivity to oxidative stress of such cells to wild-type levels (42). Thus, we used the same BAC clone to create an *FXN* genomic reporter system by inserting a luciferase sequence in exon 5a of *FXN*, generating pBAC-*FXN-Luc* vector. This was performed using a selection/counter-selection homologous recombination protocol based on RecET recombination, which allows precise modification of BAC vectors without leaving behind unwanted sequences such as bacterial antibiotic resistance cassettes, which can affect gene expression (53). pBAC-*FXN-Luc* vector encodes an FXN-luciferase fusion protein of the size of ∼79 kDa expressed under the *FXN* native promoter, combining all the advantages of a reporter system with the physiologically relevant expression levels obtained by using the full *FXN* genomic DNA locus. pBAC-*FXN-Luc* vector carries normal allele GAA repeats (six GAA) in intron 1; hence, we replaced these six GAA repeats with an ∼310 GAA repeat expansion generating pBAC-*FXN-GAA-Luc*. We demonstrate by sequencing and *Mbo*II digestion that the GAA expansion consists of pure GAA repeats and does not contain interruptions.

When generating stable clonal cell lines, random vector integration could affect *FXN* transgene expression levels considerably, masking differences in *FXN* expression caused by expanded GAA repeats. For this reason, we used site-specific integration to allow controlled comparison of the two vectors. We adapted a well-described site-specific integration protocol to the BAC technology by modifying pBAC-*FXN-Luc* and pBAC-*FXN-GAA-Luc* vectors to carry a promoter-less FRT-hygromycin cassette. In the event of random vector integration in the genome, these vectors are unable to provide hygromycin resistance; however, when co-transfected with an Flp recombinase-encoding plasmid in an acceptor cell line carrying a CMV-FRT-LacZeo construct, site-specific integration is achieved and the hygromycin cassette is expressed providing resistant clonal cell lines. FISH and copy number analysis are necessary for the generation of comparable cell lines, with one copy of the transgene and sharing the same integration site within the genome. To generate our cell model, we selected HEK-293 for their high efficiency of transfection with BAC vectors and for their ease of propagation, which makes them ideal for high-throughput screenings. We isolated two reporter clonal cell lines, *FXN-Luc* and *FXN-GAA-Luc*, and we demonstrate that the presence of ∼310 GAA repeats causes a reduction in FXN-luciferase mRNA and protein levels via a heterochromatin-mediated silencing mechanism. We show that the GAA expansion induces a decrease in histone acetylation and an increase in histone methylation at the promoter and at the regions upstream and downstream of the GAA repeats, in accordance with previous reports (26,27,32). Furthermore, *FXN-GAA-Luc* cells show increased CpG methylation upstream and downstream of the GAA repeat expansion when compared with *FXN-Luc* cells. Upstream of the GAA repeats, we observe increased DNA methylation at CpG sites 4 and 5, in accordance with previous data obtained using post-mortem brain and heart tissues from FRDA patients (27) and peripheral blood and buccal cells from FRDA patients (29). Our results for CpG site 6 are different from methylation data from FRDA patients (27,28) but are in agreement with data from brain tissue of FRDA mice (27). In the region downstream of the GAA repeats, we observed increased methylation at CpG sites 1, 2 and 7, whereas the rest of the CpG sites were completely methylated. Unlike our findings, analysis of methylation on cells (29) and tissues (27) from patients reported a general hypomethylation in this region. A methylation pattern similar to our findings in this region has been reported in YG8 mice, which do not present hypomethylation in this region (27). Since our results are different from the data obtained from patients, who carry large GAA expansions, but are similar to the findings in YG8 mice which carry shorter GAA repeats (190 + 90 GAA repeats), we speculate that this

change in CpG methylation could be related to the size of the GAA expansion. To explain how shorter repeat expansions cause changes in DNA methylation, Al-Mahdawi *et al.* (27) suggested that large GAA expansions increase the distance of the downstream region from a putative methylation centre at the 5′ end of the Alu sequence, located upstream of GAA repeats, thus generating hypomethylation of the region downstream of GAA repeats when compared with control individuals. However, a shorter GAA expansion would decrease such distance, thereby leading to hypermethylation in this region.

To our knowledge, this is the first GAA-expanded genomic reporter system that allows direct comparison of the effect of normal or expanded *FXN* loci on *FXN* expression. A previous *FXN* genomic reporter system has been described; however, this system does not contain expanded GAA repeats (54,55). The endogenous *FXN* gene is present in our cell model; however, detection of expression by luciferase assay and qRT-PCR targeted to the exon 5a–luciferase junction allow specific detection of transgene expression from the *FXN-Luc* and *FXN-GAA-Luc* vectors.

Cell models represent a very important tool for understanding the pathological repression of expression resulting from expanded GAA repeats and currently there are no such reporter cell models which allow direct comparison of normal and GAA-expanded genomic DNA loci. A few reporter cell models have been described so far which carry long expanded GAA repeats; however, these models do not allow the study of such repeats on *FXN* expression, since they typically carry only long GAA repeats within a reporter gene lacking any *FXN* gene sequence (31,39) or only contain a small portion of intron 1 (38). Thus, although they provide a suitable model to study GAA instability, previous models do not provide a tool to analyze how *FXN* expression is reduced in the presence of GAA repeats. Genomic DNA vectors, instead, provide the full genomic locus of a gene of interest, including the native promoter, all introns and exons and all elements necessary for regulated physiological transgene expression.

Currently, there is no proven treatment for FRDA. Future therapies are likely to be represented by combination of therapies targeting the reduced *FXN* expression, the increased oxidative stress and the intra-mitochondrial iron accumulation (36,55). However, drugs targeting the reduced *FXN* expression are the most promising strategies, since they target the primary molecular defect of FRDA. Recently, a few promising molecules have been proposed which increase *FXN* expression in patient cells (26,34,35,55). However, to increase the likelihood of developing successful treatments, it is important to increase the efficiency of discovery of such molecules. Reporter models of FRDA which reproduce the GAA-mediated *FXN* repression and the epigenetic hallmarks of FRDA represent an excellent tool to achieve this. We believe that our cell model fits these criteria and represents an excellent platform for the screening of *FXN* up-regulating compounds, since it allows high-throughput analysis of chemical libraries due to the rapid detection of FXN-luciferase expression by luciferase assay.

Here we show how the efficiency of screening can be increased by performing a pre-selection of small molecules based on their structure, since this considerably reduces the size of the primary screening. Our compound screening resulted in the identification of a novel compound which increases *FXN* expression in FRDA patient cells by increasing histone acetylation around the GAA repeats. This finding strengthens the hypothesis that formation of highly packaged heterochromatin is involved in the transcriptional silencing induced by GAA repeats (26). C5 is a novel small molecule which shows promising potential as a therapy for FRDA as we demonstrate it achieves a consistent 1.5–2-fold *FXN* up-regulation on our cell model, on untransfected HEK cells and most importantly on primary cells from three different FRDA patients. We show in a direct side-by-side comparison a significantly greater up-regulating effect of C5 on the expanded *FXN* locus compared with the wild-type locus, demonstrating a relative specificity for GAA expansions. However, we do observe a significant increase in endogenous *FXN* expression in non-mutant HEK cells. To explain these observations, we could hypothesize that C5 affects *FXN* expression through two separate mechanisms of action, one affecting both wild-type and expanded loci and one specific to the expanded *FXN* locus. This would cause an increase in *FXN* expression in the wild-type *FXN* locus but generate a greater increase in the expanded locus. Alternatively, we could hypothesize a single mechanism of action for C5 but a different responsiveness of the expanded and the wild-type *FXN* loci, generating a different extent of up-regulation in response to C5. Regarding the extent of the *FXN* up-regulation and the specificity for the expanded *FXN* locus, C5 is comparable with other *FXN*-increasing compounds currently undergoing clinical trials, which makes C5 a promising compound. The majority of these molecules are not characterized by specificity for the expanded *FXN* locus: PPAR-γ agonists induce an ∼1.86-fold increase in *FXN* mRNA in wild-type fibroblasts and an ∼2-fold increase in FRDA primary fibroblasts (56); resveratrol elevates frataxin mRNA by 2-fold in FRDA cells and increases frataxin protein both in wild-type cells and in FRDA cells (55); erythropoietin increases frataxin protein by 1.5-fold in control individuals and by 2-fold in FRDA patients (57); nicotinamide, a class III HDAC inhibitor, up-regulates *FXN* expression in primary lymphocytes from both healthy individuals and FRDA patients, with a higher increase in expression on FRDA samples (58). The pimelic *o*-aminobenzamide HDAC inhibitors show specificity for the pathogenic *FXN* allele. The latest of such compounds show a similar or higher up-regulation of *FXN* mRNA in primary lymphocytes from patients than C5 (34,59); however, the first HDAC inhibitor identified, BML-210, induced a 2-fold up-regulation in *FXN* expression, which is comparable with the elevation induced by C5 (26).

It is worth noting that *FXN* up-regulation is the most important requirement for developing a therapy for FRDA and we do not require that a therapy for FRDA should be specific to the GAA-expanded locus. FRDA patients do not carry a wild-type allele and this implies that a compound which can successfully increase *FXN* expression but which shows less specificity for the GAA repeats is as promising as a specific one. Finally, we believe that the most likely therapeutic approach for FRDA will be a combination of therapies aimed at targeting the frataxin deficit through multiple mechanisms of action. The combined use of unspecific *FXN*-increasing compounds with molecules acting specifically on the GAA-mediated silencing may be the most promising approach to increase frataxin protein levels to those observed in asymptomatic carriers.

HDAC inhibitors are currently considered the most promising therapy for FRDA and our results suggest C5 might belong to this class of molecules, although further studies are required to identify the exact mechanism of action.

In conclusion, we have generated a novel *FXN*-reporter expression cell model which recapitulates the *FXN* gene repression seen in FRDA patients and provides a versatile tool for the dissection of the mechanism causing the *FXN* transcriptional silencing induced by expanded GAA repeats. We demonstrate one application of this tool which resulted in the identification of a promising small molecule for the therapy of FRDA. Future studies will focus on chemical modifications of C5 to generate analogues with improved activity and chemical properties.

MATERIALS AND METHODS

Ethics statement

Informed consent for participation in the study was obtained according to the Declaration of Helsinki and approved by the Central Oxford Research Ethics Committee and the Research and Development Department of the Oxford Radcliffe Hospitals NHS Trust, Oxford.

Vector construction

The BAC clone RP11-265B8 carrying the whole 80 kb *FXN* locus with exons 1 to 5b of the *FXN* gene (42) was used to generate pBAC-*FXN-Luc*. To insert the GSGSG-luciferase sequence in exon 5a, we used a selection/counter-selection homologous recombination protocol based on RedET recombination (GeneBridges, Dresden, Germany). The recombination using the RPSL-Neo and the pSC101-BAD plasmids was carried out in two steps in *Escherichia coli* according to the manufacturer's instructions. In the first step, a PCR product containing the RPSL-Neo cassette flanked by 58 bp homology arms to either side of exon 5a was used to insert the cassette in exon 5a. In the second step, the RPSL-Neo cassette was replaced with a PCR product containing the GSGSG-luciferase sequence flanked by ~155 bp homology arms. The luciferase sequence was obtained by PCR amplification on pGL3-promoter vector (Promega). Successful construction was confirmed by PCR and sequencing. To replace the ~6 GAA repeats present in pBAC-*FXN-Luc* with ~310 GAA repeats, the RPSL-Neo cassette was amplified with primers carrying 58 bp homology arms to sequences immediately upstream and downstream of GAA repeats and the product was inserted in intron 1 of pBAC-*FXN-Luc*. Subsequently, the RPSL-Neo was replaced with a PCR product containing 280 and 830 GAA repeats amplified from NA16207 using GAA-F and GAA-R primers (1). Owing to low recombination efficiency, successful recombinant bacterial colonies were identified by colony blot assay and later confirmed by Southern blot. Colony blot and Southern blot were carried out using DIG High Prime DNA Labeling and Detection Starter Kit II (Roche), according to the manufacturer's instructions and using a digoxigenin (DIG)-labelled TTC_{10} probe for detection. Sequencing of GAA repeats on the TTC strand was performed using 602R. Since sequencing of the GAA strand is blocked by the presence of a long stretch of A immediately upstream of the GAA repeats, we designed a primer called LX-1st-GAA(5′-TACTAAAAAATACAAAAAAAAAAAA-AAAAGAAG-3′) which overcomes this area and allows sequencing of the GAA strand. To test the purity of the GAA expansion, we performed a PCR of pBAC-*FXN-Luc* and pBAC-*FXN-GAA-Luc* vectors using primers GAA-F and GAA-R (1) and digested the products with *Mbo*II, according to a method described by Holloway *et al*. (46). *Cre/loxP*-mediated retrofitting of pBAC-*FXN-Luc* and pBAC-*FXN-GAA-Luc* vectors to iBAC vector and to pH-FRT-Hy was performed as previously described (40). pH-FRT-Hy was obtained by modifying pcDNA5/FRT (Life Technologies) (61). Packaging into HSV-1 amplicons was carried out as previously described (47,48).

Cell culture and primary lymphocytes

HEK-293 cells were cultured in DMEM medium supplemented with 10% fetal bovine serum (FBS), 2 mM L-glutamine, 100 U/ml penicillin/streptomycin. HEK FRT cells were generated by transfecting the pFRT-*lacZeo* plasmid (Life Technologies), followed by selection in medium containing 100 μg/ml Zeocin (Life Technologies). *FXN-GAA-Luc* and *FXN-Luc* clonal cell lines were propagated in complete DMEM medium (see above) supplemented with 100 μg/ml Hygromycin B (Life Technologies). For the CMV-Luc experiment, HEK cells were seeded in a 24-well plate at a density of 2×10^5 cells/well, and after 24 h, they were transfected with a CMV-Luc-expressing plasmid using Lipofectamine (Life Technologies) and Plus Reagent (Life Technologies). Epstein–Barr virus-transformed lymphoblastoid cell lines GM15850 (from individuals affected by FRDA, alleles with 1030 and 650 GAA repeats) and GM15851 (from an unaffected sibling with normal range of GAA repeats) were obtained from the Human Genetic Cell Repository of the Coriell Institute (USA) and propagated in RPMI1640 medium supplemented with 15% FBS and 2 mM L-glutamine. SH-SY5Y cells were cultured in DMEM/F-12 supplemented with 10% FBS, 2 mM L-glutamine, 100 U/ml penicillin/streptomycin. SH-SY5Y cells were infected with iBAC-*FXN-Luc* and iBAC-*FXN-GAA-Luc* vectors packaged in HSV-1 amplicons as previously reported (40,48). Blood was collected from anonymous individuals affected by FRDA using Vacutainer tubes. Only patients who were shown to have two expanded alleles using a PCR-based assay were included in the study. The assays were performed by accredited molecular genetic testing laboratories in the UK. Primary lymphocytes were isolated from blood samples by centrifugation on Ficoll-Paque PLUS gradient (GE Healthcare) in Leucosep tubes (Greiner Bio-One), according to the protocol described by Miltenyl Biotec. Cell viability was monitored using trypan blue exclusion and purified cells maintained in RPMI1640 medium supplemented with 15% FBS and 2 mM L-glutamine. Compounds were added to the purified cells at the concentrations indicated above and incubated for 72 h. Cells were spun down at 200*g* for 5 min at room temperature and lysed for RNA total extraction. All cells were grown at 37°C in 5% CO_2.

PCR, qPCR and copy number assay

Genomic DNA from *FXN-GAA-Luc* and *FXN-Luc* clonal cell lines was isolated by standard phenol/chloroform extraction and ethanol precipitation. PCR amplification of the GAA

repeat sequence was carried out on 200 ng of genomic DNA using primers 147F and 602R (61) and Expand Long Template DNA Polymerase (Roche) as previously described (1,45). Vector integration at the docking site was assessed by PCR analysis using primers pSV40-F (5′-CCAGTTCCGCCCATTCTC-3′) and Hygro-R (5′-CAGCTATTTACCCGCAGGAC-3′) using AmpliTaq Gold (Roche). For copy number determination, genomic DNA from *FXN-Luc* and *FXN-GAA-Luc* clonal cell lines was isolated using Illustra Tissue and Cells GenomicPrep Mini Spin Kit (GE Healthcare), according to the manufacturer's instructions. The number of transgene copies was determined by real-time PCR, using the relative standard curve method. Five-fold dilution standards were prepared to generate a standard curve for each primer pair. The upstream region of the GAA repeats in intron 1 of the *FXN* gene was amplified using the primers UpGAA-F and UpGAA-R (26) and normalizing data by GAPDH, using the primers GAPDH-F and GAPDH-R (26). Three independent genomic DNA samples per cell line were quantified in triplicate by real-time PCR using the SYBR Green PCR Master Mix (Applied Biosystems). Each 25 μl reaction contained 2 μl of genomic DNA dilution, 1 × SYBR Green PCR Master Mix and 70 nM of each primer. The assay was performed using the StepOnePlus Real-Time PCR system (Applied Biosystems) with the following protocol: 10 min at 95°C for enzyme activation, followed by 40 cycles of denaturation at 95°C for 15 s and primer annealing and extension at 60°C for 1 min. Specificity of amplification was monitored with a final dissociation stage which generates a melting curve. The number of transgene copies was determined by comparison with the acceptor cell line HEK FRT as reference sample, which carries three endogenous *FXN* loci as determined by FISH.

qRT-PCR

Total RNA from HEK FRT, *FXN-Luc* and *FXN-GAA-Luc* clonal cell lines was extracted using RNeasy Mini Kit (Qiagen) and treated with RNase-Free DNase (Qiagen). cDNA was synthesized from 1 μg of total RNA using random primers (Life Technologies) and SuperScript III Reverse Transcriptase (Life Technologies) in a reaction volume of 20 μl. qPCR was carried out as described above, using qFXN-Luc-F (5′-CGG AAAAGATGCTGGAAGTG-3′) and qFXN-Luc-R (5′-AACC AGGGCGTATCTCTTCA-3′) for FXN-luciferase mRNA detection, FXN-F and FXN-R (26) for *FXN* mRNA detection, and data were normalized to GAPDH. Total RNA from primary lymphocytes was extracted and treated with DNase using RNAqueous-Micro Kit (Life Technologies). RNA was reverse-transcribed as above. *FXN* mRNA was detected using FXN-F and FXN-R (26) and data normalized to GAPDH, HPRT and Beta-Actin using the following primers: GAPDH-2F (5′-GGTCTCCTCTGACTTCAACA-3′) and GAPDH-2R (5′-AGCCAAATTCGTTGTCATAC-3′) (RTPrimer DB, ID: 912), HPRT-F (5′-GCCAGACTTTGTTGGATTTG-3′) and HPRT-R (5′-CTCTCATCTTAGGCTTTGTATTTTG-3′) (RTPrimerDB, ID: 984), ACTB-F (5′-AGCGCGGCTACAGCTTCA-3′) and ACTB-R (5′-CGTAGCACAGCTTCTCCTTAATGTC-3′) (RTPrimerDB, ID: 2203).

Luciferase assay

FXN-GAA-Luc and *FXN-Luc* clonal cell lines were counted with trypan blue or Scepter (Millipore) and seeded in 6 cm dishes (1.5×10^6 cells/dish), 24-well (1×10^5 cells/well) or 96-well (3×10^4 cells/well) format. When assaying luciferase expression, cells were washed with PBS and lysed in Lysis Buffer (25 mM Tris–PO_4, pH7.8, 2 mM CDTA, 10% glycerol and 1% Triton X-100) for 20 min at 4°C. Seventy-five microlitres of lysates were mixed with 100 μl of Luciferase Assay Buffer (15 mM $MgSO_4$, 15 mM KPO_4, pH 7.8, 4 mM EGTA, pH 7.8, 2 mM ATP and 2 mM DTT) and 50 μl of D-Luciferin (0.3 mg/ml). The relative light units of luciferase of each cell line were determined using the Dynex MLX 96 Well Plate Luminometer and were normalized by total protein concentration, determined using Bicinchoninic acid solution (BCA, Sigma).

Compounds library and screening

The complete library is made of 25 000 compounds dissolved in DMSO at a concentration of 2.5 mg/ml. For the primary screening, *FXN-GAA-Luc* cells were seeded in 96 wells at a density of 3×10^4 cells/well and incubated in triplicate with the 88 preselected compounds at a final concentration of 20 μM for 48 h. Luciferase assay was performed as described above. An authentic sample of C5 [1-(3,4-dimethylphenoxy)-3-(4-morpholinyl)-2-propanol hydrochloride] was purchased from ChemBridge Corporation (ID: 5358626); *m/z* (ESI+) 266 (100%, $[M+H]^+$), 288 (35%, $[M+Na]^+$).

Chromatin immunoprecipitation

Analysis of histone modifications in the *FXN* promoter and regions flanking GAA repeats in lymphoblastoid (GM15850 and GM15851), *FXN-Luc* and *FXN-GAA-Luc* cell lines was performed as previously described (62). Lymphoblastoid cell lines were treated with DMSO or C5 for 48 h before cross-linking. Proteins were cross-linked to DNA by 1% formaldehyde treatment for 4 min (GM15850 and GM15851) or 7 min (*FXN-Luc* and *FXN-GAA-Luc* cells). Chromatin from lysed cells was sheared by sonication to obtain fragments from 100 to 800 bp using a Bioruptor (Diagenode). Immunoprecipitation experiments were performed using the Immunoprecipitation Kit with Dynabeads Protein G (Life Technologies), according to the manufacturer's instructions. Chromatin was incubated with one of the following antibodies at 4°C: (i) anti-H3K9ac (07–352), (ii) anti-H4K8ac (07–328), (iii) anti-H3K9me2 (07–441), (v) anti-H3K9me3 (07–442) and (vi) normal rabbit serum (12–370) used as a negative control (all antibodies were purchased from Millipore). Samples were treated with 500 mM NaCl, 0.25 mg/ml RNase A and 0.25 mg/ml proteinase K and incubated at 65°C overnight to reverse cross-linking. After reversing the cross-linking, the amount of *FXN* DNA immunoprecipitated was quantified in triplicate by real-time PCR using the SYBR Green PCR Master Mix (Applied Biosystems) and determined using the $\Delta\Delta C_t$ method. The immunoprecipitated DNA was normalized to 10% of input and taking into consideration the background signal. For the analysis of the *FXN* promoter and the regions upstream and downstream of the GAA repeats, we used primers described in Herman *et al*. (26).

Bisulfite sequencing

Genomic DNA of the *FXN-Luc* and *FXN-GAA-Luc* cell lines was isolated using the Illustra Tissue and Cells GenomicPrep Mini Spin Kit (GE Healthcare). Two micrograms of DNA was used in the bisulfite conversion reaction using the EpiTect Bisulfite kit (Qiagen), according to the manufacturer's instructions. Nested PCR was performed on bisulfite-converted DNA using HotStart Taq DNA Polymerase (Qiagen) with the following primers, described in Al-Mahdawi *et al*. (27): F1G and R1G (first-round PCR) and F2G and R2G (second-round PCR) for the upstream region of the GAA repeats in intron 1 of the *FXN* gene; NH1F and SLGR2 (first-round PCR) and NH2F and SLGR1 (second-round PCR) for the downstream region of the GAA repeats in intron 1 of the *FXN* gene. PCR products were cloned into pGEM-T easy vector. A total of 10 colonies were sequenced per region and for each cell line.

Analysis of frataxin protein

For western blot analysis of FXN-luciferase protein, *FXN-Luc* and *FXN-GAA-Luc* cells were washed in PBS and lysed in RIPA buffer (50 mM, Tris, pH 8, 150 mM NaCl, 2 mM EGTA, 0.5% sodium deoxycholate, 1% Igepal 630, 0.1% sodium dodecyl sulphate (SDS)] with protease inhibitors (Complete Mini, EDTA-free, Roche). Cell disruption was performed by repeated pipetting followed by sonication on ice (1.5 s for 10 times) (Misonix XL-2000 sonicator). Cell lysates were spun at 1300*g* for 15 min at 4°C, supernatant transferred to a new tube and protein concentration was determined through BCA assay. Protein samples were reduced in Laemmli buffer and incubated for 5 min at 100°C. Fifty micrograms of protein was resolved on 10% SDS–polyacrylamide gel electrophoresis (SDS–PAGE). Following transfer on a PVDF-type membrane (Immobilon P, Millipore), protein samples were incubated with the following antibodies: mouse monoclonal anti-luciferase (Santa Cruz, sc-57604, 1/1000 dilution) and rabbit polyclonal anti-Beta-Actin (Abcam, ab8227, 1/1000 dilution). Analysis of frataxin protein in HEK cells was carried out using Frataxin Dipstick Assay Kit (Mitosciences), according to the manufacturer's instructions. Protein concentration in cell lysates was quantified by BCA assay and 10 μg of total protein was loaded in each well. Dried dipsticks were imaged with Chemidoc XRS system (Bio-Rad) and quantification was performed using the Image J software.

Fluorescence *in situ* hybridization

Chromosome preparation and FISH analyses were carried out as previously described (63) using the unmodified *FXN* BAC and the plasmid pH-FRT-Hy as probes. Transgene integration on chromosome 1 was confirmed using a chromosome 1 centromeric probe.

ACKNOWLEDGEMENTS

We would like to thank Dr Emanuela Volpi and her core group at the Wellcome Trust Centre for Human Genetics for FISH analysis on clonal cell lines. Dr Volpi's work was supported by the Wellcome Trust (Wellcome Trust Core Award, grant number 090532/Z/09/Z). We would also like to thank R. Valentine for organizing samples from FRDA patients, H. Storr for help with mass spectrometry, T. Roberts for advice on real-time PCR, C. Coxon for help with purification of primary lymphocytes, J. Knight for advice and suggestions and A. Velayos-Baeza for the generous gift of reagents. We are particularly grateful to the patients and their families for participating in this study.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by Ataxia UK (#7125 to M.M.P.L.), the Friedreich's Ataxia Research Alliance (FARA) (to M.M.P.L. and to R.W.-M.), BabelFAmily (to M.M.P.L.), Associazione Italiana per la lotta alle Sindromi Atassiche (AISA) (to M.M.P.L.), the European Union 7th Framework Program EFACTS (grant agreement No. 242193) (to R.W.-M.), National Ataxia Foundation (NAF) (to R.W.-M.), Association Française de l'Ataxie de Friedreich (AFAF) (to R.W.-M.) and the Friedreich Ataxia Research Association (Australasia) (FARA(A)) (to M.M.P.L.). A generous donation for equipment funding was provided by C. Yvette. R.W.-M. was a Wellcome Trust Research Career Development Fellow. M.M.P.L. is an Ataxia UK Research Fellow and is currently co-funded by FARA and AISA. A.M.S. is supported by Fundação para a Ciência ea Tecnologia PhD studentship (SFRH/BD/61048/2009). J.A.-A. is supported by the Monument Trust Discovery Award from Parkinson's UK. A.H.N. is supported by DeNDRoN.

REFERENCES

1. Campuzano, V., Montermini, L., Molto, M.D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A. *et al.* (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science*, **271**, 1423–1427.
2. Schulz, J.B., Boesch, S., Bürk, K., Dürr, A., Giunti, P., Mariotti, C., Pousset, F., Schöls, L., Vankan, P. and Pandolfo, M. (2009) Diagnosis and treatment of Friedreich ataxia: a European perspective. *Nat. Rev. Neurol.*, **5**, 222–234.
3. Tsou, A.Y., Paulsen, E.K., Lagedrost, S.J., Perlman, S.L., Mathews, K.D., Wilmot, G.R., Ravina, B., Koeppen, A.H. and Lynch, D.R. (2011) Mortality in Friedreich ataxia. *J. Neurol. Sci.*, **307**, 46–49.
4. Koeppen, A.H., Michael, S.C., Knutson, M.D., Haile, D.J., Qian, J., Levi, S., Santambrogio, P., Garrick, M.D. and Lamarche, J.B. (2007) The dentate nucleus in Friedreich's ataxia: the role of iron-responsive proteins. *Acta Neuropathol.*, **114**, 163–173.
5. Koeppen, A.H., Morral, J.A., Davis, A.N., Qian, J., Petrocine, S.V., Knutson, M.D., Gibson, W.M., Cusack, M.J. and Li, D. (2009) The dorsal root ganglion in Friedreich's ataxia. *Acta Neuropathol.*, **118**, 763–776.
6. Delatycki, M.B., Williamson, R. and Forrest, S.M. (2000) Friedreich ataxia: an overview. *J. Med. Genet.*, **37**, 1–8.
7. Cossée, M., Dürr, A., Schmitt, M., Dahl, N., Trouillas, P., Allinson, P., Kostrzewa, M., Nivelon-Chevallier, A., Gustavson, K.H., Kohlschütter, A. *et al.* (1999) Friedreich's ataxia: point mutations and clinical presentation of compound heterozygotes. *Ann. Neurol.*, **45**, 200–206.
8. Cossée, M., Puccio, H., Gansmuller, A., Koutnikova, H., Dierich, A., LeMeur, M., Fischbeck, K., Dollé, P. and Koenig, M. (2000) Inactivation of the Friedreich ataxia mouse gene leads to early embryonic lethality without iron accumulation. *Hum. Mol. Genet.*, **9**, 1219–1226.
9. Filla, A., De Michele, G., Cavalcanti, F., Pianese, L., Monticelli, A., Campanella, G. and Cocozza, S. (1996) The relationship between trinucleotide (GAA) repeat length and clinical features in Friedreich ataxia. *Am. J. Hum. Genet.*, **59**, 554–560.
10. Montermini, L., Richter, A., Morgan, K., Justice, C.M., Julien, D., Castellotti, B., Mercier, J., Poirier, J., Capozzoli, F., Bouchard, J.P. *et al.*

(1997) Phenotypic variability in Friedreich ataxia: role of the associated GAA triplet repeat expansion. *Ann. Neurol.*, **41**, 675–682.
11. Schmucker, S. and Puccio, H. (2010) Understanding the molecular mechanisms of Friedreich's ataxia to develop therapeutic approaches. *Hum. Mol. Genet.*, **19**, R103–R110.
12. Puccio, H., Simon, D., Cossée, M., Criqui-Filipe, P., Tiziano, F., Melki, J., Hindelang, C., Matyas, R., Rustin, P. and Koenig, M. (2001) Mouse models for Friedreich ataxia exhibit cardiomyopathy, sensory nerve defect and Fe-S enzyme deficiency followed by intramitochondrial iron deposits. *Nat. Genet.*, **27**, 181–186.
13. Martelli, A., Wattenhofer-Donzé, M., Schmucker, S., Bouvet, S., Reutenauer, L. and Puccio, H. (2007) Frataxin is essential for extramitochondrial Fe-S cluster proteins in mammalian tissues. *Hum. Mol. Genet.*, **16**, 2651–2658.
14. Babcock, M., de Silva, D., Oaks, R., Davis-Kaplan, S., Jiralerspong, S., Montermini, L., Pandolfo, M. and Kaplan, J. (1997) Regulation of mitochondrial iron accumulation by Yfh1p, a putative homolog of frataxin. *Science*, **276**, 1709–1712.
15. Rötig, A., de Lonlay, P., Chretien, D., Foury, F., Koenig, M., Sidi, D., Munnich, A. and Rustin, P. (1997) Aconitase and mitochondrial iron-sulphur protein deficiency in Friedreich ataxia. *Nat. Genet.*, **17**, 215–217.
16. Wong, A., Yang, J., Cavadini, P., Gellera, C., Lonnerdal, B., Taroni, F. and Cortopassi, G. (1999) The Friedreich's ataxia mutation confers cellular sensitivity to oxidant stress which is rescued by chelators of iron and calcium and inhibitors of apoptosis. *Hum. Mol. Genet.*, **8**, 425–430.
17. Wells, R.D. (2008) DNA triplexes and Friedreich ataxia. *FASEB J.*, **22**, 1625–1634.
18. Bidichandani, S.I., Ashizawa, T. and Patel, P.I. (1998) The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. *Am. J. Hum. Genet.*, **62**, 111–121.
19. Sakamoto, N., Chastain, P.D., Parniewski, P., Ohshima, K., Pandolfo, M., Griffith, J.D. and Wells, R.D. (1999) Sticky DNA: self-association properties of long GAA.TTC repeats in R.R.Y triplex structures from Friedreich's ataxia. *Mol. Cell*, **3**, 465–475.
20. Sakamoto, N., Ohshima, K., Montermini, L., Pandolfo, M. and Wells, R.D. (2001) Sticky DNA, a self-associated complex formed at long GAA*TTC repeats in intron 1 of the frataxin gene, inhibits transcription. *J. Biol. Chem.*, **276**, 27171–27177.
21. Grabczyk, E., Mancuso, M. and Sammarco, M.C. (2007) A persistent RNA.DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP *in vitro*. *Nucleic Acids Res.*, **35**, 5351–5359.
22. Vetcher, A.A., Napierala, M., Iyer, R.R., Chastain, P.D., Griffith, J.D. and Wells, R.D. (2002) Sticky DNA, a long GAA.GAA.TTC triplex that is formed intramolecularly, in the sequence of intron 1 of the frataxin gene. *J. Biol. Chem.*, **277**, 39217–39227.
23. Ohshima, K., Montermini, L., Wells, R.D. and Pandolfo, M. (1998) Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication *in vivo*. *J. Biol. Chem.*, **273**, 14588–14595.
24. Grabczyk, E. and Usdin, K. (2000) The GAA*TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. *Nucleic Acids Res.*, **28**, 2815–2822.
25. Saveliev, A., Everett, C., Sharpe, T., Webster, Z. and Festenstein, R. (2003) DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing. *Nature*, **422**, 909–913.
26. Herman, D., Jenssen, K., Burnett, R., Soragni, E., Perlman, S.L. and Gottesfeld, J.M. (2006) Histone deacetylase inhibitors reverse gene silencing in Friedreich's ataxia. *Nat. Chem. Biol.*, **2**, 551–558.
27. Al-Mahdawi, S., Pinto, R.M., Ismail, O., Varshney, D., Lymperi, S., Sandi, C., Trabzuni, D. and Pook, M. (2008) The Friedreich ataxia GAA repeat expansion mutation induces comparable epigenetic changes in human and transgenic mouse brain and heart tissues. *Hum. Mol. Genet.*, **17**, 735–746.
28. Greene, E., Mahishi, L., Entezam, A., Kumari, D. and Usdin, K. (2007) Repeat-induced epigenetic changes in intron 1 of the frataxin gene and its consequences in Friedreich ataxia. *Nucleic Acids Res.*, **35**, 3383–3390.
29. Evans-Galea, M.V., Carrodus, N., Rowley, S.M., Corben, L.A., Tai, G., Saffery, R., Galati, J.C., Wong, N.C., Craig, J.M., Lynch, D.R. *et al.* (2012) FXN methylation predicts expression and clinical outcome in Friedreich ataxia. *Ann. Neurol.*, **71**, 487–497.
30. Castaldo, I., Pinelli, M., Monticelli, A., Acquaviva, F., Giacchetti, M., Filla, A., Sacchetti, S., Keller, S., Avvedimento, V.E., Chiariotti, L. *et al.* (2008) DNA methylation in intron 1 of the frataxin gene is related to GAA repeat length and age of onset in Friedreich ataxia patients. *J. Med. Genet.*, **45**, 808–812.
31. Soragni, E., Herman, D., Dent, S.Y., Gottesfeld, J.M., Wells, R.D. and Napierala, M. (2008) Long intronic GAA*TTC repeats induce epigenetic changes and reporter gene silencing in a molecular model of Friedreich ataxia. *Nucleic Acids Res.*, **36**, 6056–6065.
32. Kumari, D., Biacsi, R.E. and Usdin, K. (2011) Repeat expansion affects both transcription initiation and elongation in Friedreich ataxia cells. *J. Biol. Chem.*, **286**, 4209–4215.
33. De Biase, I., Chutake, Y.K., Rindler, P.M. and Bidichandani, S.I. (2009) Epigenetic silencing in Friedreich ataxia is associated with depletion of CTCF (CCCTC-binding factor) and antisense transcription. *PLoS One*, **4**, e7914.
34. Rai, M., Soragni, E., Jenssen, K., Burnett, R., Herman, D., Coppola, G., Geschwind, D.H., Gottesfeld, J.M. and Pandolfo, M. (2008) HDAC inhibitors correct frataxin deficiency in a Friedreich ataxia mouse model. *PLoS One*, **3**, e1958.
35. Sturm, B., Stupphann, D., Kaun, C., Boesch, S., Schranzhofer, M., Wojta, J., Goldenberg, H. and Scheiber-Mojdehkar, B. (2005) Recombinant human erythropoietin: effects on frataxin expression *in vitro*. *Eur. J. Clin. Invest.*, **35**, 711–717.
36. Wilson, R.B. (2012) Therapeutic developments in Friedreich ataxia. *J. Child. Neurol.*, **27**, 1212–1216.
37. Velasco-Sánchez, D., Aracil, A., Montero, R., Mas, A., Jiménez, L., O'Callaghan, M., Tondo, M., Capdevila, A., Blanch, J., Artuch, R. *et al.* (2011) Combined therapy with idebenone and deferiprone in patients with Friedreich's ataxia. *Cerebellum*, **10**, 1–8.
38. Grant, L., Sun, J., Xu, H., Subramony, S.H., Chaires, J.B. and Hebert, M.D. (2006) Rational selection of small molecules that increase transcription through the GAA repeats found in Friedreich's ataxia. *FEBS Lett.*, **580**, 5399–5405.
39. Ditch, S., Sammarco, M.C., Banerjee, A. and Grabczyk, E. (2009) Progressive GAA.TTC repeat expansion in human cell lines. *PLoS Genet.*, **5**, e1000704.
40. Lufino, M.M., Manservigi, R. and Wade-Martins, R. (2007) An S/MAR-based infectious episomal genomic DNA expression vector provides long-term regulated functional complementation of LDLR deficiency. *Nucleic Acids Res.*, **35**, e98.
41. Wade-Martins, R., White, R.E., Kimura, H., Cook, P.R. and James, M.R. (2000) Stable correction of a genetic deficiency in human cells by an episome carrying a 115 kb genomic transgene. *Nat. Biotechnol.*, **18**, 1311–1314.
42. Gomez-Sebastian, S., Gimenez-Cassina, A., Diaz-Nido, J., Lim, F. and Wade-Martins, R. (2007) Infectious delivery and expression of a 135 kb human FRDA genomic DNA locus complements Friedreich's ataxia deficiency in human cells. *Mol. Ther.*, **15**, 248–254.
43. Puccio, H. and Koenig, M. (2000) Recent advances in the molecular pathogenesis of Friedreich ataxia. *Hum. Mol. Genet.*, **9**, 887–892.
44. Robinson, C.R. and Sauer, R.T. (1998) Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. *Proc. Natl Acad. Sci. USA*, **95**, 5929–5934.
45. Al-Mahdawi, S., Pinto, R.M., Ruddle, P., Carroll, C., Webster, Z. and Pook, M. (2004) GAA repeat instability in Friedreich ataxia YAC transgenic mice. *Genomics*, **84**, 301–310.
46. Holloway, T.P., Rowley, S.M., Delatycki, M.B. and Sarsero, J.P. (2011) Detection of interruptions in the GAA trinucleotide repeat expansion in the FXN gene of Friedreich ataxia. *Biotechniques*, **50**, 182–186.
47. Wade-Martins, R., Smith, E.R., Tyminski, E., Chiocca, E.A. and Saeki, Y. (2001) An infectious transfer and expression system for genomic DNA loci in human and mouse cells. *Nat. Biotechnol.*, **19**, 1067–1070.
48. Saeki, Y., Fraefel, C., Ichikawa, T., Breakefield, X.O. and Chiocca, E.A. (2001) Improved helper virus-free packaging system for HSV amplicon vectors using an ICP27-deleted, oversized HSV-1 DNA in a bacterial artificial chromosome. *Mol. Ther.*, **3**, 591–601.
49. Westwood, I.M., Bhakta, S., Russell, A.J., Fullam, E., Anderton, M.C., Kawamura, A., Mulvaney, A.W., Vickers, R.J., Bhowruth, V., Besra, G.S. *et al.* (2010) Identification of arylamine *N*-acetyltransferase inhibitors as an approach towards novel anti-tuberculars. *Protein Cell*, **1**, 82–95.
50. Russell, A.J., Westwood, I.M., Crawford, M.H., Robinson, J., Kawamura, A., Redfield, C., Laurieri, N., Lowe, E.D., Davies, S.G. and Sim, E. (2009) Selective small molecule inhibitors of the potential breast cancer marker,

human arylamine *N*-acetyltransferase 1, and its murine homologue, mouse arylamine *N*-acetyltransferase 2. *Bioorg. Med. Chem.*, **17**, 905–918.
51. Rai, M., Soragni, E., Chou, C.J., Barnes, G., Jones, S., Rusche, J.R., Gottesfeld, J.M. and Pandolfo, M. (2010) Two new pimelic diphenylamide HDAC inhibitors induce sustained frataxin upregulation in cells from Friedreich's ataxia patients and in a mouse model. *PLoS One*, **5**, e8825.
52. Gimenez-Cassina, A., Wade-Martins, R., Gomez-Sebastian, S., Corona, J.C., Lim, F. and Diaz-Nido, J. (2011) Infectious delivery and long-term persistence of transgene expression in the brain by a 135-kb iBAC-FXN genomic DNA expression vector. *Gene Ther.*, **18**, 1015–1019.
53. Zhang, Y., Muyrers, J.P., Testa, G. and Stewart, A.F. (2000) DNA cloning by homologous recombination in *Escherichia coli*. *Nat. Biotechnol.*, **18**, 1314–1317.
54. Sarsero, J.P., Li, L., Wardan, H., Sitte, K., Williamson, R. and Ioannou, P.A. (2003) Upregulation of expression from the FRDA genomic locus for the therapy of Friedreich ataxia. *J. Gene. Med.*, **5**, 72–81.
55. Li, L., Voullaire, L., Sandi, C., Pook, M.A., Ioannou, P.A., Delatycki, M.B. and Sarsero, J.P. (2013) Pharmacological screening using an FXN-EGFP cellular genomic reporter assay for the therapy of Friedreich ataxia. *PLoS One*, **8**, e55940.
56. Marmolino, D., Acquaviva, F., Pinelli, M., Monticelli, A., Castaldo, I., Filla, A. and Cocozza, S. (2009) PPAR-gamma agonist Azelaoyl PAF increases frataxin protein and mRNA expression: new implications for the Friedreich's ataxia therapy. *Cerebellum*, **8**, 98–103.
57. Acquaviva, F., Castaldo, I., Filla, A., Giacchetti, M., Marmolino, D., Monticelli, A., Pinelli, M., Sacca, F. and Cocozza, S. (2008) Recombinant human erythropoietin increases frataxin protein expression without increasing mRNA expression. *Cerebellum*, **7**, 360–365.
58. Chan, P.K., Torres, R., Yandim, C., Law, P.P., Khadayate, S., Mauri, M., Grosan, C., Chapman-Rothe, N., Giunti, P., Pook, M. *et al.* (2013) Heterochromatinization induced by GAA-repeat hyperexpansion in Friedreich's ataxia can be reduced upon HDAC inhibition by vitamin B3. *Hum. Mol. Genet.*, **22**, 2662–2675.
59. Soragni, E., Xu, C., Plasterer, H.L., Jacques, V., Rusche, J.R. and Gottesfeld, J.M. (2012) Rationale for the development of 2-aminobenzamide histone deacetylase inhibitors as therapeutics for Friedreich ataxia. *J. Child Neurol.*, **27**, 1164–1173.
60. Alegre-Abarrategui, J., Christian, H., Lufino, M.M.P., Mutihac, R., Venda, L.L., Ansorge, O. and Wade-Martins, R. (2009) LRRK2 regulates autophagic activity and localizes to specific membrane microdomains in a novel human genomic reporter cellular model. *Hum. Mol. Genet.*, **18**, 4022–4034.
61. Montermini, L., Andermann, E., Labuda, M., Richter, A., Pandolfo, M., Cavalcanti, F., Pianese, L., Iodice, L., Farina, G., Monticelli, A. *et al.* (1997) The Friedreich ataxia GAA triplet repeat: premutation and normal alleles. *Hum. Mol. Genet.*, **6**, 1261–1266.
62. Carey, M.F., Peterson, C.L. and Smale, S.T. (2009) Chromatin immunoprecipitation (ChIP). *Cold Spring Harb. Protoc.*, **4**, 1–8.
63. Jefferson, A. and Volpi, E.V. (2010) Fluorescence *in situ* hybridization (FISH) for genomic investigations in rat. *Methods Mol. Biol.*, **659**, 409–426.

ANNEX 2

Silva AM, Brown JM, Buckle VJ, Wade-Martins R and Lufino MM (2015) Expanded GAA Repeats Impair *FXN* Gene Expression and Reposition the *FXN* Locus to the Nuclear Lamina in Single Cells. *Hum. Mol. Genet.* 24(12):3457-71.

Human Molecular Genetics, 2015, 1–15

doi: 10.1093/hmg/ddv096
Advance Access Publication Date: 26 March 2015
Original Article

ORIGINAL ARTICLE

Expanded GAA repeats impair *FXN* gene expression and reposition the *FXN* locus to the nuclear lamina in single cells

Ana M. Silva[1,2], Jill M. Brown[3], Veronica J. Buckle[3], Richard Wade-Martins[1,*] and Michele M.P. Lufino[1,*]

[1]Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford OX1 3QX, UK, [2]Faculdade de Medicina, Universidade de Lisboa, Lisboa 1649-028, Portugal and [3]Medical Research Council, Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK

*To whom correspondence should be addressed. Tel: +44 1865282837; Fax: +44 1865272420; Email: richard.wade-martins@dpag.ox.ac.uk (R.W.-M.); Tel: +44 1865282668; Fax: +44 1865272420; Email: michele.lufino@dpag.ox.ac.uk (M.M.P.L.)

Abstract

Abnormally expanded DNA repeats are associated with several neurodegenerative diseases. In Friedreich's ataxia (FRDA), expanded GAA repeats in intron 1 of the frataxin gene (*FXN*) reduce *FXN* mRNA levels in averaged cell samples through a poorly understood mechanism. By visualizing *FXN* expression and nuclear localization in single cells, we show that GAA-expanded repeats decrease the number of *FXN* mRNA molecules, slow transcription, and increase *FXN* localization at the nuclear lamina (NL). Restoring histone acetylation reverses NL positioning. Expanded GAA-*FXN* loci in FRDA patient cells show increased NL localization with increased silencing of alleles and reduced transcription from alleles positioned peripherally. We also demonstrate inefficiencies in transcription initiation and elongation from the expanded GAA-*FXN* locus at single-cell resolution. We suggest that repressive epigenetic modifications at the expanded GAA-*FXN* locus may lead to NL relocation, where further repression may occur.

Introduction

An abnormal GAA trinucleotide repeat expansion in intron 1 of the frataxin gene (*FXN*) causes Friedreich's ataxia (FRDA; OMIM 229300), a progressive neurodegenerative disease and the most common form of recessive ataxia (1). *FXN* alleles in healthy individuals contain <36 GAA repeats, whereas in FRDA patients GAA expansions ranging from 70 to 1700 GAA repeats lead to *FXN* mRNA deficiency and subsequent reduced levels of frataxin, a nuclear-encoded mitochondrial protein essential for life (1,2). GAA expansion-mediated transcriptional dysregulation occurs due to the generation of unusual DNA structures such as triplexes or sticky DNA (3,4), R-loops (5,6) and heterochromatin (7,8), which lead to increased DNA methylation at specific CpG sites (9–11), reduced histone acetylation (H3/H4ac) and increased levels of methylated histones H3K9me2 and H3K9me3 (8,10). It has been suggested that these epigenetic changes surrounding the GAA expansion impair RNA polymerase II (RNAPII) elongation (12), but also spread upstream towards the *FXN* promoter, inducing a non-permissive chromatin configuration for transcription initiation, altering nucleosome positioning and preventing binding of insulator CCCTC-binding factor (CTCF) (6,13–15). However, these studies only provide the probable state of the *FXN* gene, as these observations come from experiments in which the outputs of bulk cell cultures are averaged. A dissection of the silencing mechanism in FRDA *in situ*, in which *FXN* localization and expression are quantified at single-cell level, is crucial to improve our understanding of the underlying pathogenesis and ultimately to design effective therapies for FRDA.

Received: January 13, 2015. **Revised and Accepted:** March 12, 2015

Studies indicate that the radial positioning of a gene within the nucleus correlates with its transcriptional output, but whether a gene is transcribed due to its position or its position is determined by its transcriptional state is the subject of current research (16–19). In particular, genomic DNA interactions with the nuclear periphery (NP) can actively contribute to gene repression (20–22). However, this is not a general phenomenon (20,23), rather is gene-specific, and may depend on multiple parameters such as transcription factor accessibility, promoter strength, existence of insulator elements and pre-existing chromatin marks, which may counteract the mechanisms underlying transcription repression. The nuclear lamina (NL) tends to be in contact with heterochromatin and is associated with markers of gene repression, such as enrichment in histone modifications H3K9me2 and H3K27me3 and depletion of activating histone marks and RNAPII occupancy (reviewed in ref. 24). Given that the expanded *FXN* heterochromatic state is coupled with gene repression, we asked where GAA-expanded *FXN* alleles are found in the nucleus, and how their location impacts on repression.

Here, we report a single-cell analysis of *FXN* repression in which we identify the NL as a novel and key player in *FXN* transcriptional impairment and silencing. Using a multidisciplinary approach including analysis in both fixed and living single cells, we show that expanded GAA repeats increase *FXN* positioning at the NL, leading to decreased numbers of *FXN* mRNA molecules and slower transcription kinetics in an *FXN-GAA-MS2* cell model. We observe the same abnormal repositioning to the NL in carrier and FRDA patient cells and show that this tightly correlates with a marked decrease in the number of actively expressing *FXN* alleles. Furthermore, we show that those few active expanded *FXN* alleles located at the NL express at a significantly lower level than the alleles located in the interior of the nucleus. Finally, we demonstrate that expanded GAA repeats predominantly disrupt *FXN* transcription initiation. The mechanisms we describe may extend to other genetic diseases mediated by repeat expansions within regions of non-coding DNA.

Results

GAA repeat expansion increases *FXN-GAA-MS2* positioning at the NL

To investigate the link between *FXN* localization and expression at the single-cell level, we modified our previously described reporter model (25), which carries the whole 80 kb *FXN* locus with its native promoter, including exons 1–5b and all regulatory elements necessary to achieve physiologically relevant *FXN* expression, a luciferase reporter gene in exon 5a and the presence of either six GAA repeats or a ~310 GAA repeat expansion in intron 1. To fluorescently label the transgenic *FXN* mRNA and quantify the effect of the expansion on *FXN* transcription, we inserted 24 repeats of MS2 protein-binding sites (MBS) into exon 2 by homologous recombination, generating pBAC-*FXN-MS2-Luc* and pBAC-*FXN-GAA-MS2-Luc* (Fig. 1A). We used site-specific integration to generate several stable human clones of *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cell lines carrying a single copy of each vector in the same integration site in chromosome 1 (25) (Supplementary Material, Fig. S1A–D), because this allows a direct single-allele comparison of normal and expanded *FXN* loci. We selected *FXN-MS2-Luc* clone 2 and *FXN-GAA-MS2-Luc* clone 22 (referred to throughout the text as *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc*, respectively) for the following studies, as these best represented the average reduction in *FXN* mRNA (Supplementary Material, Fig. S1E) and protein levels (Supplementary Material, Fig. S1F) in screened clones. The ~310 GAA repeat expansion in the *FXN-GAA-MS2-Luc* cell line recapitulates the characteristic *FXN* gene repression seen in FRDA by altering the *FXN* epigenetic landscape (25) and reducing *FXN* mRNA levels by 62% and frataxin protein levels by 52% (Fig. 1B).

FXN transgene localization in *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* lines was determined by immunofluorescence *in situ* hybridization (Immuno-FISH), using co-localized probes for the vector backbone and *FXN* loci to distinguish the transgene from the three additional endogenous *FXN* loci (25) present in each clone (Fig. 1C). We found *FXN* at the NL in ~44% of *FXN-GAA-MS2-Luc* cells compared to only ~10% of *FXN-MS2-Luc* cells. Treatment with the histone deacetylase inhibitors (HDACi) nicotinamide and compound 106—previously shown to upregulate *FXN* expression and increase histone acetylation levels and DNAse I accessibility at the expanded *FXN* locus in FRDA patient cells (26,27)—repositioned only *FXN-GAA-MS2-Luc* away from the NL (Fig. 1D), but increased transgene *FXN* mRNA levels in both cell lines (Fig. 1E), suggesting a complex interplay between *FXN* repression and intranuclear localization.

GAA-expanded repeats decrease the number of *FXN* mRNA molecules and slow transcription kinetics in single *FXN-GAA-MS2* cells

To further understand *FXN* repression, we analysed the transcriptional output of individual transgenic *FXN* alleles in *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells by RNA FISH using a probe hybridizing to the MBS sequence. Single *FXN-MS2-Luc* mRNA molecules were identified as sharp signals above non-specific fluorescence HEK cell background in accordance with previous studies demonstrating single-molecule resolution of this technique (Supplementary Material, Fig. S2) (28–30). The insertion of the MBS in exon 2 allowed us to identify all fluorescence spots as mature *FXN* mRNA molecules, considering transcription sites were not distinguishable. *FXN-GAA-MS2-Luc* cells contained ~5 ± 2 mRNAs per cell, and *FXN-MS2-Luc* cells contained ~9 ± 4 mRNA per cell (Fig. 2A), therefore ~310 GAA repeats reduce the number of mature mRNA molecules by 44% at single-cell level. This quantification is in agreement with oligonucleotide array assays previously showing the majority of mRNAs are present at <10 copies per cell (31).

We then followed the kinetic behaviour of *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* transcription using fluorescence recovery after photobleaching (FRAP) (30). Nascent transcripts were detected as fluorescent spots as they were bound by MS2-YFP fusion protein and visualized in real time (Supplementary Material, Video S1 and S2). Recovery of fluorescence indicates new transcripts are actively synthesized at the transcription site (Fig. 2B). FRAP curves showed that the time required to reach steady state after photobleaching was different between the two cell lines. *FXN-MS2-Luc* cells showed half-time of recovery ($t_{1/2}$) = 12.90 s with full recovery = 120 s (Fig. 2C and Supplementary Material, Video S3), whereas *FXN-GAA-MS2-Luc* cells showed slower kinetics $t_{1/2}$ = 48.10 s with full recovery = 260 s (Fig. 2C and Supplementary Material, Video S4). Since we found no difference in *FXN* mRNA stability in *FXN-GAA-MS2-Luc* and *FXN-GAA-MS2-Luc* cells (Fig. 2D) (12), RNA FISH and FRAP data indicate that the GAA expansion decreases the amount of *FXN-GAA-MS2-Luc* mRNA molecules by impeding RNAPII transcription initiation and/or elongation. We could not distinguish between the two events given the placement of the MBS in exon 2, where the FRAP readout is obtained.

To elucidate the link between *FXN* localization and repression, we analysed the fluorescence intensities of active transgenes in living *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells. The

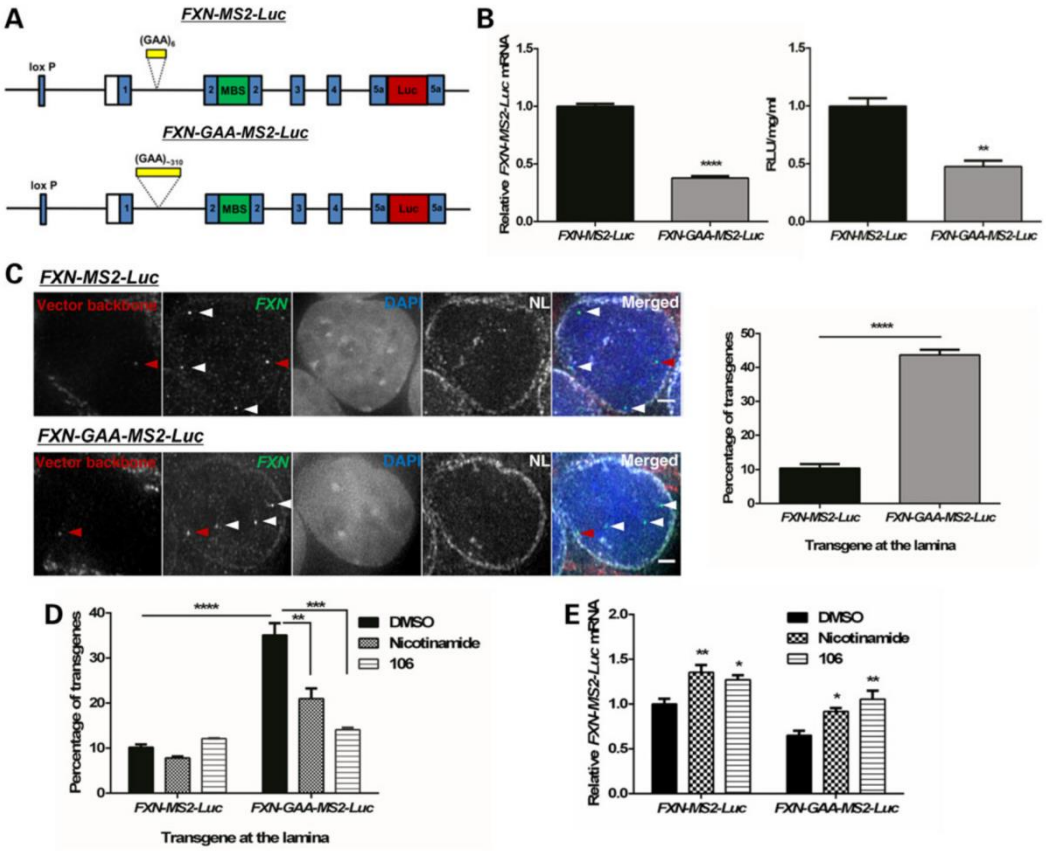

Figure 1. The expanded GAA repeat *FXN* transgene associates with the NL more frequently in an *FXN-GAA-MS2-Luc* cell model. (**A**) Schematic representation of the pBAC-*FXN-MS2-Luc* and pBAC-*FXN-GAA-MS2-Luc* vectors. Each vector carries either six or ~310 GAA repeats in intron 1, an array of 24 MBS in exon 2 and expresses an FXN-luciferase fusion protein. (**B**) *FXN-MS2-Luc* clone 2 and *FXN-GAA-MS2-Luc* clone 22 mRNA expression (left) was determined by qRT-PCR using qFXN-Luc primers and showed a significant reduction in *FXN-GAA-MS2-Luc* clone 22 mRNA levels, as determined by unpaired two-tailed Student's *t*-test (****$P < 0.0001$). Data are mean ± SEM ($n = 3$). Luciferase assay (right) showed a significant reduction in FXN-MS2-luciferase protein levels in *FXN-GAA-MS2-Luc* clone 22 cells, as determined by unpaired two-tailed Student's *t*-test (**$P < 0.01$). Data are mean ± SEM ($n = 3$). (**C**) 3D Immuno-FISH detection of *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* association with the NL (pseudocoloured white in merged image). Transgenes were identified by co-localization of FISH signals from *FXN* loci (green) and the vector backbone (red) in *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells (left). Nuclei were identified by DAPI staining (blue). Arrows show endogenous (white) and transgene (red) alleles. Comparison of the proportion of transgenes contacting the NL showed a significant difference in *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells (right) (****$P < 0.0001$, as determined by Fisher's two-tailed exact test, $n = 140$ cells per condition). Data are mean ± SEM from two independent FISH experiments. Association of FISH signals with the NL was analysed using ImageJ: image stacks were analysed manually, and association was determined by eye when *FXN* FISH signals were touching or superimposing the NL signal at a given z-section. Images are maximum projections of z-stacks covering the signals. Scale bar: 2 µm. (**D**) Treatment with nicotinamide (10 mM) or compound 106 (10 µM) for 48 h significantly decreased *FXN-GAA-MS2-Luc* interaction with NL (**$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$, as determined by Fisher's two-tailed exact test, $n = 140$ cells per condition). Data are mean ± SEM from 2–3 independent FISH experiments. (**E**) *FXN-MS2-Luc* mRNA expression after incubation with HDACi for 48 h was determined by qRT-PCR using qFXN-Luc primers (25). Data were normalized to *GAPDH* mRNA levels. Results are relative to *FXN-MS2-Luc* DMSO. Two-way analysis of variance (ANOVA) with Sidak's correction test showed a significant difference in the two cell lines when comparing with the DMSO control of each line (*$P < 0.05$; **$P < 0.01$). Data are mean ± SEM from three independent experiments.

fluorescence intensity of transcription sites was significantly lower when transgenes were expressed at the NP compared with the interior in both *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells (Fig. 2E). This indicates that both *FXN* transgenes express when localized at NP, although they do so at lower levels. When compared to active *FXN-MS2-Luc* transgenes, spot intensity of *FXN-MS2-GAA-Luc* transgenes was significantly lower only in the nuclear interior. Taken together with data from Figure 1C, these data suggest that expanded GAA repeats increase *FXN-GAA-MS2-Luc* positioning at the NL, where expression levels are reduced when compared to the nuclear interior or with *FXN-MS2-Luc* expression levels in the nucleoplasm.

Expanded *FXN* alleles localize more frequently at the NL in carrier cells

Reporter transgenes are a powerful tool to reveal mechanisms of gene expression, however their expression may be influenced by the local environment of the integration site (32). *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* transgenes are integrated in chromosome 1, and given that the native *FXN* is located in chromosome 9, we asked whether our findings from chromosome 1 would replicate in a more physiologically relevant context. We first analysed *FXN* localization in its native genomic environment in carrier-derived lymphoblastoid cells (GM14519) carrying 9 and 1285 GAA repeats

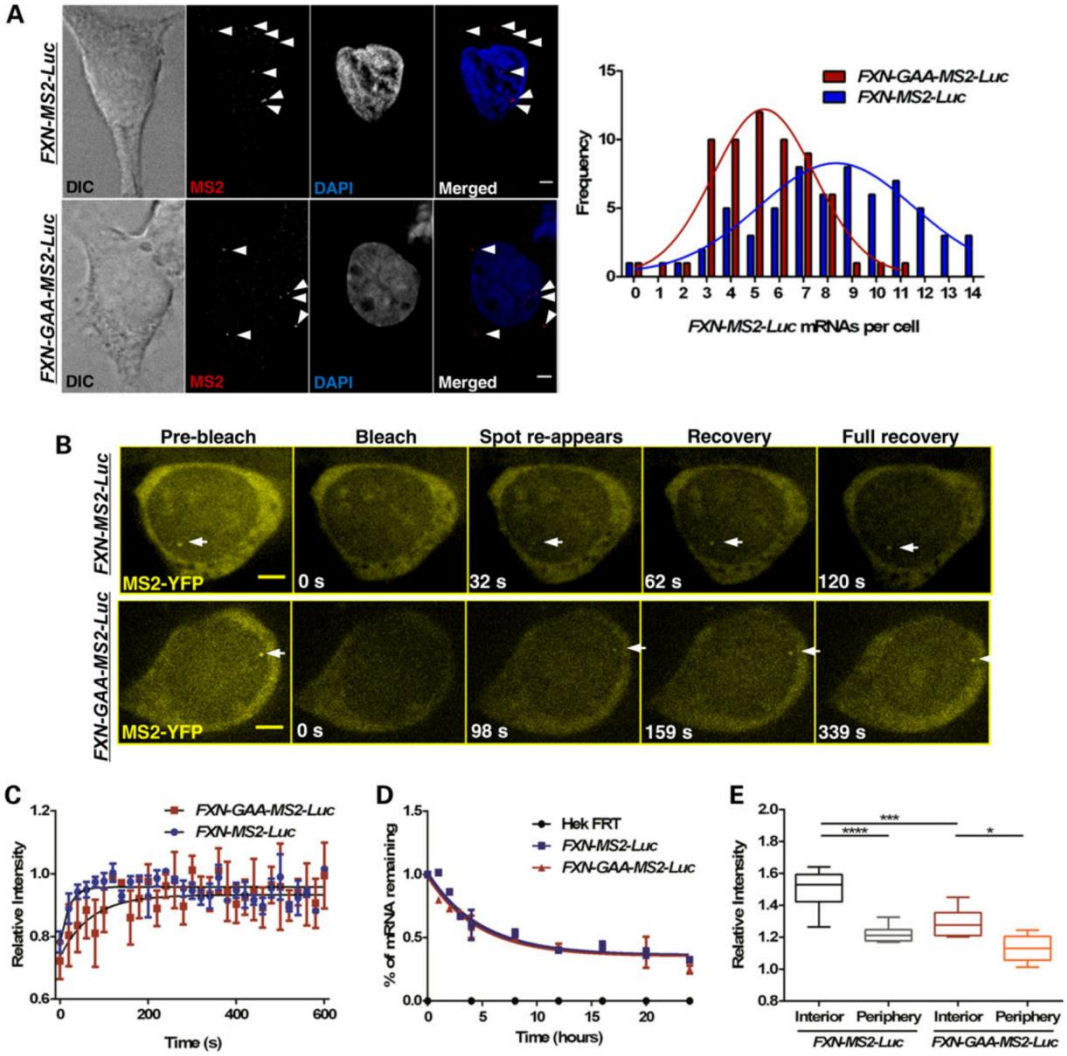

Figure 2. The GAA repeat expansion reduces *FXN-GAA-MS2-Luc* transcriptional output and impedes transcription in fixed and living *FXN-GAA-Luc* cells. (**A**) Quantification of *FXN-MS2-Luc* mRNA molecules in *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells by RNA FISH. *FXN-MS2-Luc* mRNA molecules were identified as sharp dots (arrows) above background of non-specific fluorescence using an MS2-Cy3 probe (red in merged images) in *FXN-MS2-Luc* (top images) and *FXN-GAA-MS2-Luc* (bottom images) cells. Nonlinear regression curve-fit comparison showed a significant difference between the distribution of *FXN-MS2-Luc* mRNA molecules in *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells ($P < 0.0001$, $F = 33.31$, d.f. = 3, 21, $n = 63$ cells per line) (right). Images are maximum projections of z-stacks covering the signals. (**B**) Live-cell fluorescence video frames taken before and at indicated times after photobleaching of the transcription site. Scale bar: 2 µm. (**C**) FRAP recovery curves and fitted exponential curves show a difference in transcription kinetics ($P = 0.0048$, $F = 4.434$, d.f. = 3, 210, $n = 5$ cells per line). (**D**) Degradation of *FXN-MS2-Luc* mRNA was analysed by qRT-PCR using qFXN-Luc primers in *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells after treatment with Actinomycin D for 24 h. Data were normalized to β-actin mRNA. Comparison of fitted exponential decay curves shows no difference between the two cell lines ($P = 0.5979$, $F = 0.6351$, d.f. = 3, 32). *FXN-MS2-Luc* mRNA half-life was ~4 h. Data are mean ± SEM ($n = 3$). (**E**) Quantification of fluorescence intensities of active transgenes in living *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells. Spot intensity of active *FXN* transgenes above MS2-YFP background was significantly lower in the nuclear interior in *FXN-GAA-MS2-Luc* cells when compared with *FXN-MS2-Luc* cells (***$P < 0.001$), and significantly lower when alleles were expressing at the periphery than in the nuclear interior in the two cell lines, as determined by one-way ANOVA with Sidak's multiple comparisons test (****$P < 0.0001$, *$P < 0.05$, $n = 15$ cells per line).

in each *FXN* allele, respectively. Each carrier cell exhibited a 55% reduction in mRNA compared with healthy cells (Supplementary Material, Fig. S3). Co-detection of a DNA probe for *FXN* with a $(GAA)_{15}$ or $(TTC)_{15}$ oligo signal allowed discrimination of the expanded allele and a direct intracellular comparison of the normal and the expanded *FXN* allele, with distance measurements for each allele to the NP by 2D and 3D FISH (33,34). We found that the expanded *FXN* allele localized preferentially closer to the NP than the normal allele (Fig. 3A and B). To further analyse the observed difference in nuclear positioning, we then assessed *FXN* association with specific nuclear subdomains known to either (i) suggest the transcriptional status of a gene, such as nuclear SC35-enriched speckles and chromosome territories (CT), or (ii) interact with heterochromatin, such as the nucleolus and the NL (21,35–39). *FXN* never looped out of chromosome 9 nuclear territory in 2D FISH experiments (Fig. 3C), demonstrating that this

does not account for the observed difference in nuclear localization. We found no evidence suggesting the expanded and normal *FXN* alleles associated differently with nuclear speckles (Fig. 3D) or with heterochromatin surrounding nucleoli (Fig. 3E). However, when we analysed *FXN* positioning at the NL, we found an increased number of expanded *FXN* alleles contacting with the

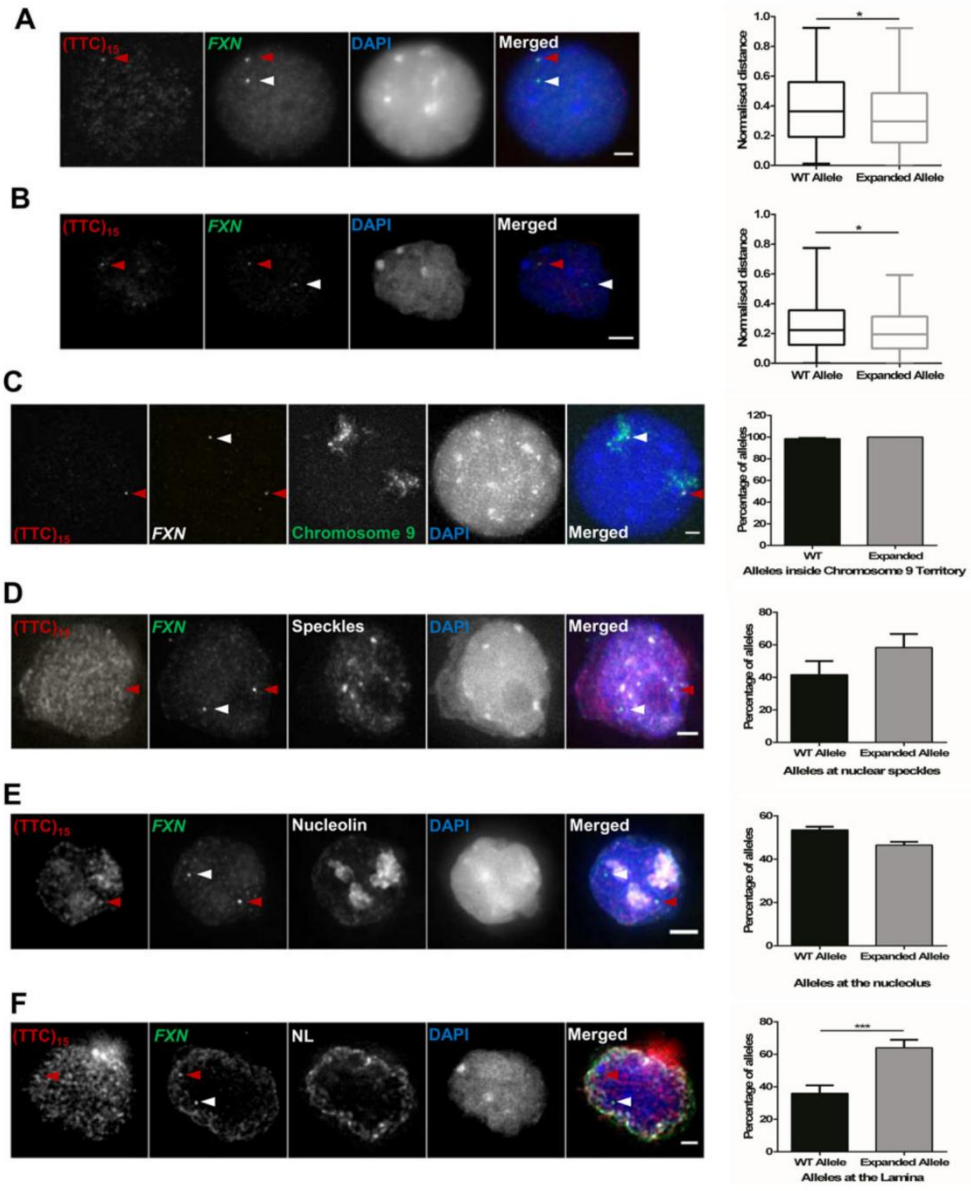

Figure 3. Expanded *FXN* allele localizes closer to the NP and sits at the NL in carrier cells (GM14519). Each allele was detected with a DNA probe for *FXN* (green in merged image) and the expanded allele was identified by co-localization with a $(GAA)_{15}$ or $(TTC)_{15}$ oligo signal (red) (left). Nuclei were identified by DAPI staining (blue). Arrows indicate normal (white) and expanded (red) alleles. (**A** and **B**) Distance measurements between FISH signals, and the NP were measured and normalized by the local ray between the nuclear centre and periphery passing through the FISH signal (right). The expanded *FXN* allele (1285 + 50 GAA repeats) localized preferentially closer to the NP than the WT (9 GAA repeats) (*$P = 0.0153$, as determined by Wilcoxon matched-paired two-tailed test; data are mean ± SEM from three independent hybridizations, $n = 147$) in 2D (**A**) and 3D DNA FISH analysis (**B**) (*$P = 0.0254$, as determined by Wilcoxon matched-paired two-tailed test; data are mean ± SEM from three independent hybridizations, $n = 107$). (**C**) *FXN* (white) did not loop out of chromosome 9 territories (green) in 2D FISH experiments. Data are mean ± SEM, $n = 136$ cells from three independent hybridizations. (**D** and **E**) 3D Immuno-FISH detection of *FXN* association with nuclear SC35-enriched speckles (D) and nucleoli (E) showed no significant difference between alleles, as determined by Fisher's two-tailed exact test (Speckles: $P = 0.4942$, $n = 86$ cells from two independent hybridizations; nucleoli: $P = 0.3607$, $n = 156$ cells from two independent hybridizations) (**F**) 3D Immuno-FISH detection of *FXN* association with the NL (pseudocoloured white). Comparison of the proportion of cells with only one allele contacting with the NL revealed the mutation induced a significant increase in co-localization with the NL, as determined by Fisher's two-tailed exact test (***$P = 0.0005$). Data are mean ± SEM from two independent hybridizations (n = 166). Images are maximum projections of z-stacks covering the signals. Association of FISH signals with subnuclear domains or chromosome territories was analysed using ImageJ: image stacks were analysed manually and the association was determined by eye when FXN FISH signals were touching or superimposing the signal from a specific subnuclear domain at a given z-section. Scale bar: 2 µm.

NL when compared with the normal allele (Fig. 3F). Taken together, these data suggest that GAA repeat expansions increase *FXN* positioning at the NL in carrier-derived cells.

NL is a key player in *FXN* transcriptional impairment and silencing

We next analysed *FXN* localization at the NL in healthy (GM15851) and patient-derived cells (GM16209) using Immuno-FISH (Fig. 4A). The patient cell line contains two *FXN* alleles carrying 800 GAA repeats each and was chosen to avoid differing behaviour of each locus due to different repeat size. Eight-hundred GAA repeats decreased mRNA levels by ~83% (Supplementary Material, Fig. S3). The distribution of cells with *FXN* FISH signals contacting with the NL was significantly different in the two cell lines (Fig. 4B). We found one or both *FXN* alleles localized at the NL in a small proportion of healthy cells (~37 and ~14%, respectively), with ~49% of cells containing both *FXN* alleles in the nucleoplasm. In contrast, the majority of FRDA cells presented both or one allele contacting with the NL (~43 and ~36%, respectively), with only 21% of FRDA cells containing both *FXN* alleles in the nucleoplasm. When data were analysed at allele level instead of cell level, ~60% of expanded *FXN* alleles were found at the NL compared with only ~33% of normal *FXN* alleles (Fig. 4B). These data suggest that the GAA expansion increases the probability of an allele to be found at the NL in FRDA cells, in agreement with data presented in Figures 1C and 3F.

To further dissect *FXN* repression, we analysed nascent *FXN* RNA synthesis at a single-cell level, using probes hybridizing with the upstream region of the GAA repeats (Fig. 4C). Figure 4D shows ~46% of healthy cells contained two *FXN* alleles expressing, and ~30 and ~24% of cells with either one or no expressing *FXN* allele, respectively. We found a higher proportion of FRDA cells (51%) containing both *FXN* alleles not expressing, followed by ~27 and ~22% of cells with either one or both *FXN* alleles expressing, respectively. Analysing these data at allele level instead of cell level, we found that ~61% of normal *FXN* alleles were active compared to only ~35% of expanded *FXN* alleles (Fig. 4D). These data suggest that the mutation increases the probability of an allele being silenced. Collectively, data from Figures 1C, 2A, E, 3F, 4B and D indicate a clear link between expanded *FXN* positioning at the NL and GAA-mediated transcriptional repression. Furthermore, these results confirm the existence of biological variability within cell populations and highlight the importance of single-cell studies (40).

To further elucidate the interplay between the repressive environment at the NL and *FXN* repression, we analysed RNA FISH spot intensities from active *FXN* alleles in Figure 4D as an indication of the quantity of transcript being synthesized at each transcription site relative to its interior or peripheral nuclear position (Fig. 4E). The signal variability of the technique was assessed by comparing RNA FISH intensities between replicates. Spot intensities were comparable among replicates (Supplementary Material, Fig. S4A), and background intensity was not significantly different in nuclei of healthy and FRDA patient cells (Supplementary Material, Fig. S4B). We found that *FXN* RNA FISH spot intensity was significantly lower when *FXN* alleles were expressed in the NP compared with the interior in both healthy and FRDA cells. This indicates that the NP was not incompatible with *FXN* expression because *FXN* RNA FISH signals could be seen at the periphery in both healthy and FRDA cells, albeit at lower intensities. Strikingly, RNA FISH spot intensity in FRDA cells was significantly lower than in healthy cells both in the interior and at the periphery. When collectively considering with the data in Figures 3F, 4B and D, we conclude that several factors combine to diminish transcription from the expanded *FXN* gene. The expanded *FXN* gene has a significantly increased likelihood of being positioned at the periphery (Figs 3F and 4B) and markedly reduced probability of being transcribed at all (Fig. 4D). Crucially, when transcription does occur at the expanded GAA-*FXN* locus, it is at a significantly reduced level both in the nucleoplasm and especially at the periphery (Fig. 4E). The NP can therefore be seen as a generally less supportive environment for *FXN* transcription, but the combined effect of expansion with peripheral relocation results in a catastrophic reduction in transcriptional output.

Expanded GAA repeats impair preferentially *FXN* transcription initiation in single FRDA cells

To analyse the effect of the mutation at the initiation and elongation stages of *FXN* transcription at single-cell resolution, probes hybridizing with the promoter after the transcription start site (TSS) and regions upstream and downstream of GAA repeats were included in our RNA FISH analysis (Fig. 5A and B). Healthy *FXN* alleles showed the expected decrease in transcription elongation throughout the analysed regions (Fig. 5C and D), confirming previous real-time reverse transcription-polymerase chain reaction (qRT-PCR) analysis (6,41). However, when compared with healthy alleles, expanded *FXN* alleles showed a ~19% decrease in RNA levels at the TSS site, demonstrating a marked reduction in *FXN* transcription initiation (Fig. 5C). We also observed a ~20% decrease at the region upstream of GAA repeats and a ~22% decrease at the downstream region of GAA repeats. We then normalized the data in Figure 5C to exclude the effect of (i) transcription initiation at the TSS and (ii) elongation through the regions upstream of GAA repeats and found that expanded *FXN* alleles showed a small but significant decrease in *FXN* mRNA synthesis only downstream and not upstream of the GAA repeat expansion when compared with the same regions in healthy alleles (Fig. 5D), demonstrating that transcription elongation through expanded GAA repeats is also affected. Taken together, these data indicate that GAA repeat expansions diminish *FXN* mRNA synthesis predominantly by impairing *FXN* transcription initiation, but also by impairing elongation through the expansion.

Discussion

Gene relocation can be driven by chromatin remodelling coupled with or without transcription changes (19). Through these experiments, we demonstrate that the GAA expansion-mediated impairment of *FXN* expression is the result of a complex interplay of multiple factors contributing to different extents to decreased *FXN* transcriptional activity (Fig. 6). Among these factors, we identify the NL, a location known to be a generally repressive environment, as a novel major player in *FXN* silencing in FRDA. Using a variety of models, we found a marked increase in expanded *FXN* positioning at the NL and a concomitant reduction in the number of active *FXN* alleles, suggesting that *FXN* localized at the NL leads to *FXN* silencing. We have also identified a small proportion of expanded *FXN* alleles sitting at the NL which are transcriptionally active, however these show significantly reduced expression when compared with alleles in the nuclear interior. We observed two distinct events affecting *FXN* expression. First, we observed that the GAA repeat expansion impairs *FXN* transcription regardless of *FXN* localization in the nuclear interior or periphery. Second, although the proportion of normal and expanded *FXN* alleles positioned at the NL is remarkably different, we found that the NL reduces the expression of normal

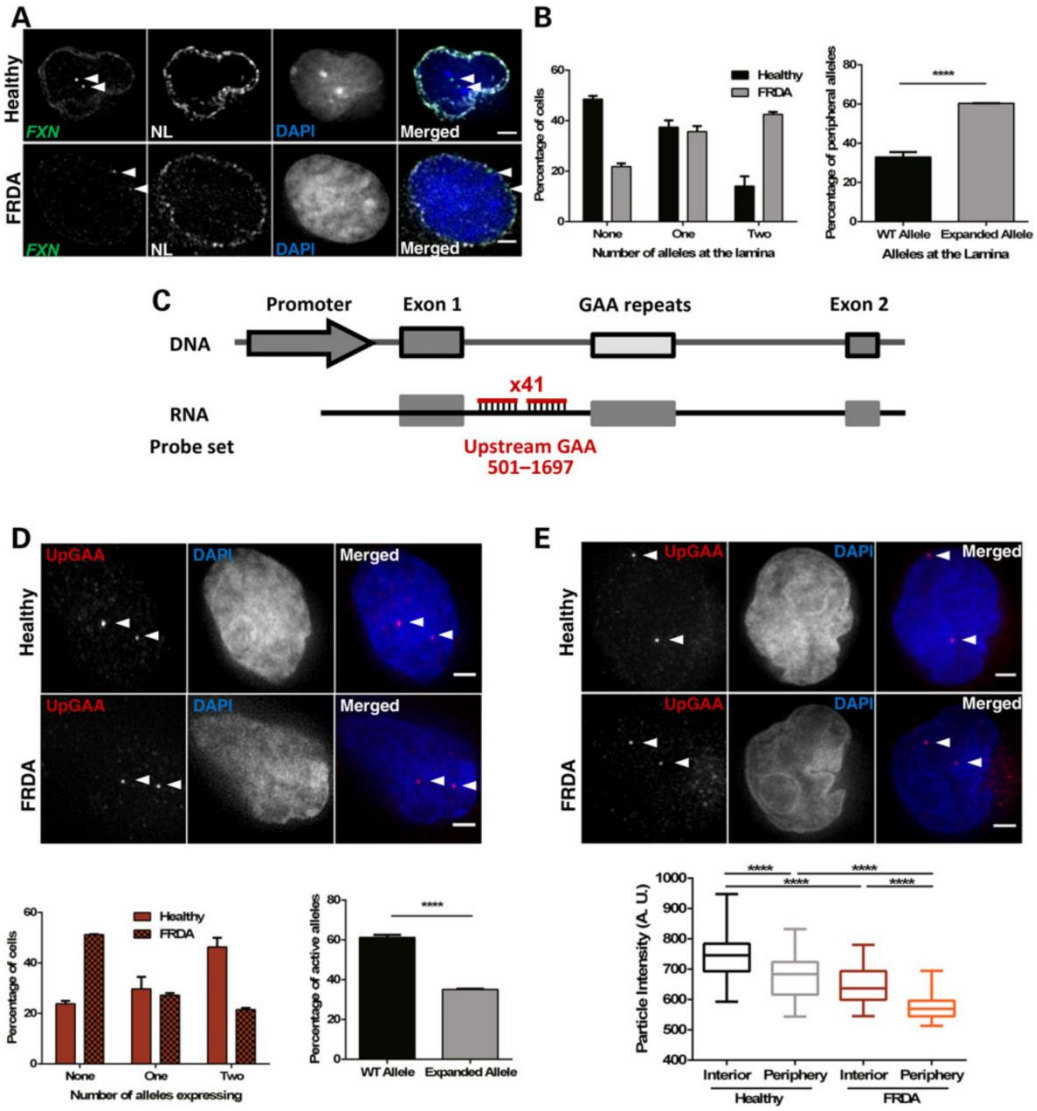

Figure 4. GAA repeat expansion increases *FXN* localization at the NL, reduces the number of active *FXN* alleles and downregulates transcription from active *FXN* in FRDA cells (GM16209). (**A**) 3D Immuno-FISH detection of *FXN* loci association with the NL in healthy (top images) and patient-derived (bottom images) cells. Each allele was detected with a DNA probe for *FXN* (green in merged image), and the NL was detected with an antibody for Lamin B1 (white). Nuclei were identified by DAPI staining (blue). Arrows indicate alleles. Association of FISH signals with the NL was analysed using ImageJ: image stacks were analysed manually and the association was determined by eye when *FXN* FISH signals were touching or superimposing the signal from the NL at a given z-section. (**B**) Histogram (left) showing the distribution of cells with *FXN* FISH signals contacting with the NL was significantly different in the two cell lines, as determined by Chi-squared test ($P < 0.0001$, $\chi^2 = 63.056$, d.f. = 2). When data were analysed at allele level instead of cell level (right), the proportion of *FXN* alleles positioned at the NL is significantly different when comparing the localization of the expanded and normal *FXN* alleles (****$P < 0.0001$, as determined by Fisher's two-tailed exact test). (**C**) Schematic representation showing the relative position to TSS and labelling of the DNA oligo probe set used in nascent RNA FISH experiments. (**D**) Nascent RNA FISH detection of *FXN* expression in healthy and patient-derived cells (top). Each allele was detected with a mix of 41 DNA oligos hybridizing with the region upstream of the GAA repeats (pseudocoloured red in merged image). Histogram (bottom left) showing the distribution of cells with RNA *FXN* FISH signals was significantly different in the two cell lines, as determined by Chi-squared test ($P < 0.0001$, $\chi^2 = 75.424$, d.f. = 2). When data were analysed at allele level instead of cell level (bottom right), the proportion of active *FXN* alleles is significantly different when comparing the expanded and normal *FXN* alleles (****$P < 0.0001$, as determined by Fisher's two-tailed exact test). Data are mean ± SEM from three independent hybridizations ($n = 187–202$ cells). (**E**) Nascent RNA FISH detection of *FXN* expression in healthy and patient-derived cells (top) showing different signal intensities of internal versus peripheral nascent RNA loci. Detection was performed as described in (D). Box plot (bottom) showing the RNA FISH spot intensity of active *FXN* alleles above the threshold was significantly lower in the interior and at the periphery in FRDA versus healthy nuclei (****$P < 0.0001$) and significantly lower when alleles were expressing at the periphery than in the nuclear interior in the two cell lines, as determined by one-way ANOVA with Tukey's multiple comparisons test (****$P < 0.0001$, $n = 245$ signals). Images are maximum projections of z-stacks covering the signals. Scale bar: 2 µm.

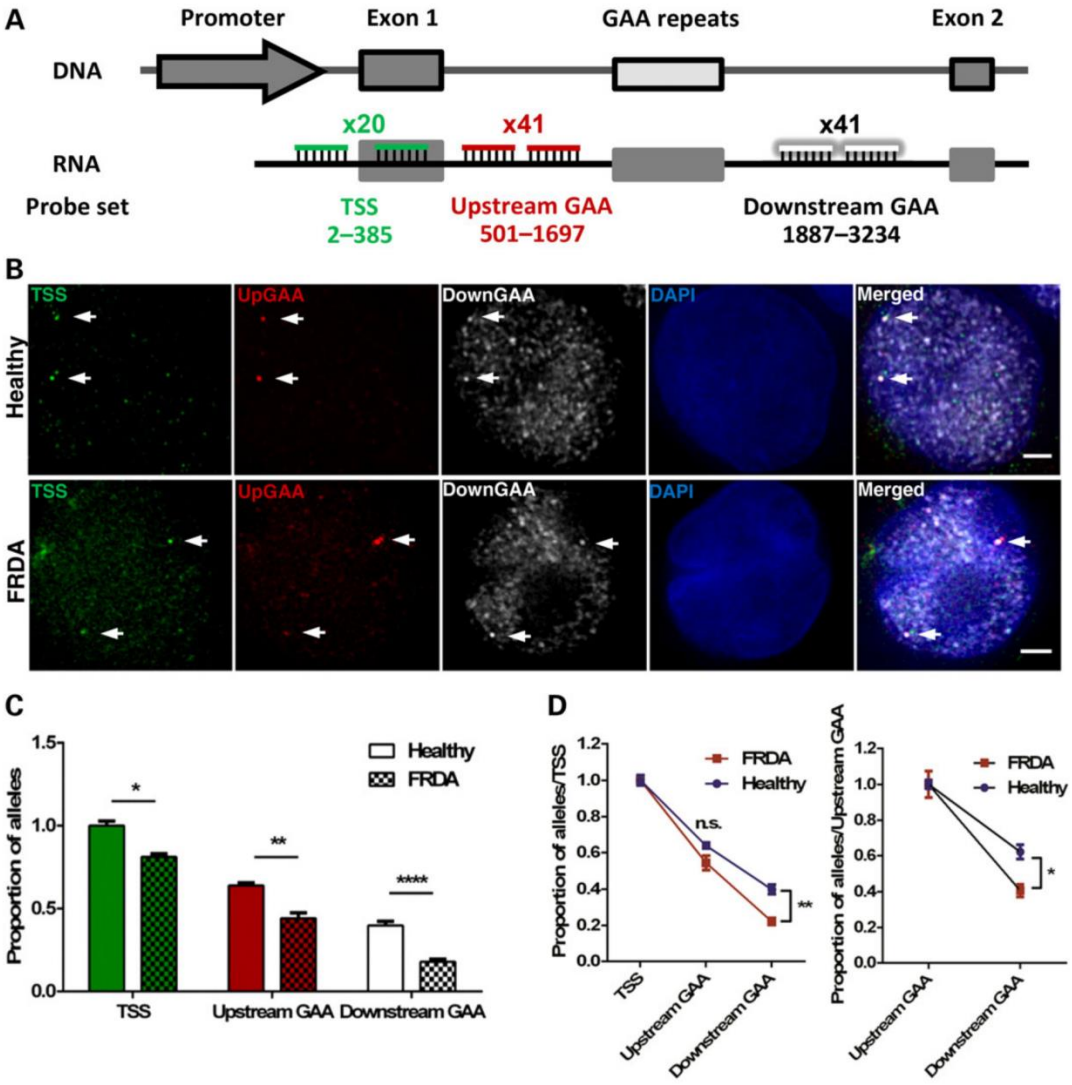

Figure 5. GAA repeat expansion disrupts transcription initiation and elongation in FRDA cells (GM16209). (**A**) Schematic representation showing the relative position to TSS and labelling of the three DNA oligo probe sets used to detect transcription initiation and elongation in RNA FISH experiments. (**B**) *FXN* RNA FISH detection of transcription initiation after the transcription starting site (TSS probe set in green) and elongation through the regions upstream (pseudocoloured red) and downstream (pseudocoloured white) of the GAA repeats. Nuclei were identified by DAPI staining (blue). Arrows indicate alleles. Images are maximum projections of z-stacks covering the signals. Scale bar: 2 µm. (**C**) Data were normalized to the TSS of healthy cells. When compared with the same region in healthy alleles, expanded *FXN* alleles showed a ∼19% decrease at the TSS, a ∼20% and a ∼22% decrease at the regions upstream and downstream of GAA repeats, respectively. This data demonstrates GAA repeat expansions reduce *FXN* transcription initiation by 19% and elongation downstream of expanded GAA repeats by 3% (*$P = 0.0119$, **$P = 0.003$, ****$P < 0.0001$, as determined by Fisher''s two-tailed exact test, $n = 218–226$ cells per line). Data are mean ± SEM from three independent FISH experiments. (**D**) Data of each line in (C) were normalized to their TSS, thus excluding the effect of transcription initiation (left) and elongation through the region upstream of GAA repeats (right). Expanded *FXN* alleles show a significant difference in the region downstream of the GAA repeats when comparing with the same region in healthy alleles (*$P < 0.05$, **$P < 0.01$, as determined by two-way ANOVA with Sidak's multiple comparisons test). Data are mean ± SEM from three independent FISH experiments.

and expanded *FXN* at the NP to a similar extent, demonstrating that the repressive action of the NL is irrespective of a GAA expansion. This leads us to a consideration of the order of molecular events occurring at the expanded *FXN* locus. One of the initial steps in *FXN* repression is the binding of nascent *FXN* mRNA to the GAA repeat expansion as the gene is being transcribed, a feature shared with the CGG repeat expansion disease Fragile X Syndrome (6,42). These RNA/DNA hybrids (R-loops) have been previously shown to form on both transgenic and endogenous expanded *FXN* loci in cell lines used in this study (6). R-loops impair RNAPII elongation and trigger epigenetic changes, such as H3K9me2 repressive marks through recruitment of G9a methyltransferase (6). The effects of inhibition of histone deacetylases were interesting in this regard. *FXN* transcription from both normal and expanded alleles is upregulated by enforced acetylation by similar amounts, yet results in significant movement from the

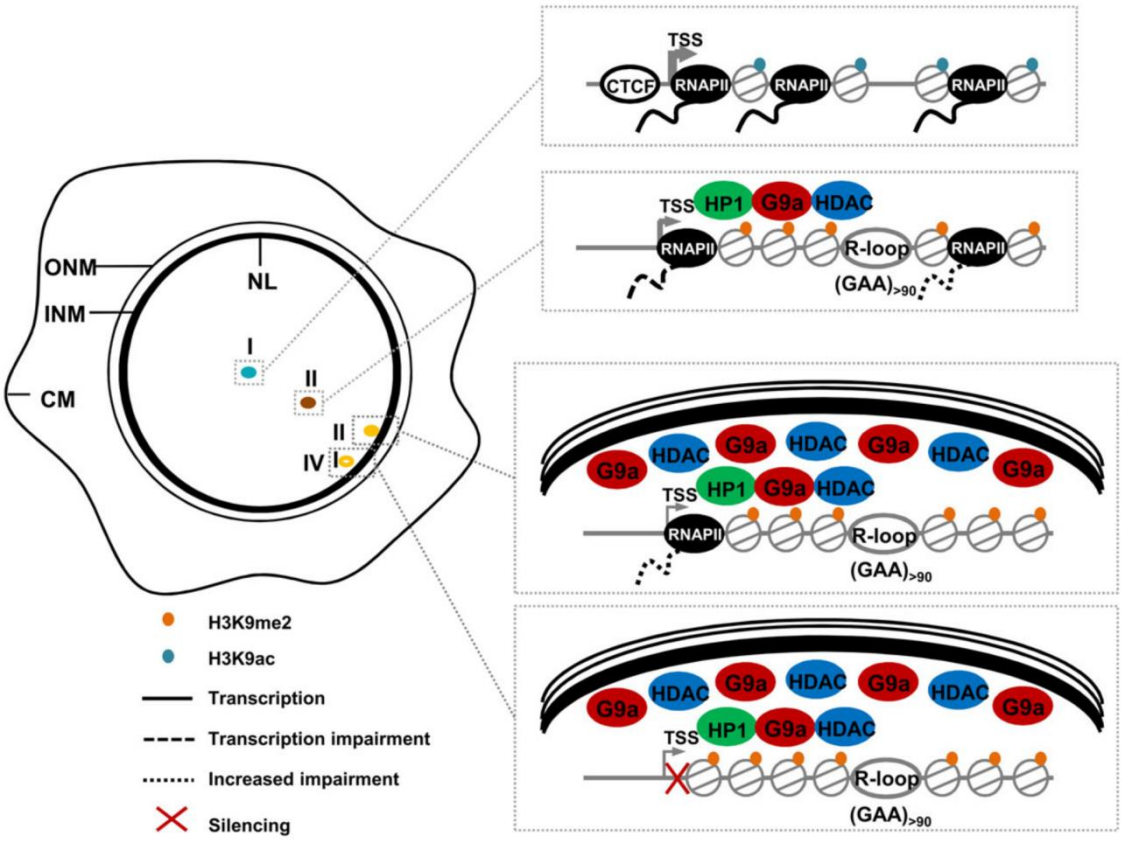

Figure 6. Model of GAA repeat-mediated *FXN* repression and the repressive role of NL in *FXN* transcriptional impairment. The euchromatic state of normal *FXN* alleles allows physiological levels of expression (I). Expanded GAA repeats trigger R-loop formation and locally recruit HP1 (13), G9a methyltransferase (6) and class I (8) and III (27) HDACs, causing epigenetic changes that contribute to transcriptional dysregulation (II). These GAA expansion-mediated structural and epigenetic changes may lead to *FXN* relocation to the NL—a location known to be a generally repressive environment and enriched in HDAC and histone methyltransferase activity—where further repression (III) and silencing (IV) occur. ONM, outer nuclear membrane; INM, inner nuclear membrane; CM, cell membrane.

NL of only the expanded alleles. The mechanism of retention at the NL of these expanded repeat structures is therefore highly sensitive to HDAC inhibition. However, the expanded alleles are unable to achieve the same level of upregulation as their normal counterparts. This could be the result of a different extent of histone acetylation achieved by normal and expanded alleles following treatment with HDAC inhibitors, which could be caused by reduced accessibility of compacted *FXN* alleles. However, the observed difference in *FXN* upregulation could also imply that other structural changes such as R-loop formation contribute to transcriptional dysregulation and that the trigger event is induced by structural and epigenetic changes to the DNA followed by an increased likelihood of retention at the NL.

It seems plausible that, once a locus adopts an abnormal DNA structure and becomes modified by specific markers, it is more likely to be retained at the NL after mitosis (43). There are several mechanisms that could be involved. Genome contacts with the NL during interphase are generally dynamic, but confined to a narrow zone near the NL. Peripheral siting of chromatin after mitosis is largely determined by a stochastic mechanism and directly linked to cellular or allelic variation in H3K9me2 levels where G9a methyltransferase acts as a regulator of NL contacts (44). The NP is enriched in heterochromatin and the presence of HP1, whose binding is promoted by methylation at lysine 9, may help to regulate the post-mitotic binding of inner nuclear membrane proteins (43). There are also spatial considerations. The *FXN* locus is located between two lamina-associated domains (LADs)—NL-interacting chromosome regions demarcated by the insulator protein CTCF that present heterochromatic histone modifications and low expression levels—in lung fibroblasts (45) and various neural cell types (46). LAD topology has yet to be assessed in the lymphoblastoid cell lines used in this study, but as different cell types present remarkable LAD similarities (46), it is tempting to speculate that *FXN* is also in an interLAD in lymphoblastoid cells. Given the severe depletion of CTCF in the 5′ UTR of the expanded *FXN* in FRDA (13), we speculate that the boundary between *FXN* and the LAD located upstream is inefficient, leading to the inclusion of expanded *FXN* inside the LADs and increased proximity to the NL, possibly through recruitment of NL-associated HDACs and G9a methyltransferase (6). It would be interesting to explore the LAD layout over this region in FRDA versus healthy individuals, and whether interruption of the boundary CTCF elements can allow consolidation of the two flanking LADs. A further spatial contribution could be that even normal *FXN* can stochastically contact the NL, considering the intranuclear mobility of chromatin (47) and the *FXN* native location on the usually peripheral chromosome 9. Indeed, we observe that transcription from the normal allele is

downregulated to an equivalent degree when located at the NP. *FXN* retention may result in a molecular interaction with the NL, as lamina-associating sequences are enriched in a GAGA motif which is bound by the transcriptional repressor cKrox associated with HDAC3 and inner nuclear membrane Lap2β (22). It would be interesting to knockdown cKrox in FRDA patient cells and analyse whether expanded *FXN* alleles at the NL relocalize to the nuclear interior. These events in turn could act through a positive feedback mechanism by mediating further impairment or silencing since the NL is enriched in HDAC and histone methyltransferase activity. Given the shared features of DNA repeat expansions, we hypothesize that the NL may play a role in heterochromatin-mediated transcriptional dysregulation in other repeat-expansion diseases and it would be interesting to interrogate the nuclear location of other repeat expansion loci. Although we show a marked degree of repression at the NP, one striking finding is that of transcriptionally active *FXN* alleles sitting at the NL, confirming what others have shown that the NP creates a repressive environment, but is not incompatible with expression (20,23). Our findings support the existence of NL heterogeneous microdomains composed of nuclear pore complexes and different lamin subtypes (24) able to impact differently on gene expression.

Non-coding triplet repeat expansions can mediate gene repression of other genes in *cis* (48,49). There is evidence suggesting that *PIP5K1B* is downregulated in FRDA lymphocytes and fibroblasts (50). The *PIP5K1B* locus is located ∼26 kb upstream of *FXN* and encodes phosphatidylinositol 4-phosphate 5-kinase β type I, an enzyme functionally linked to actin cytoskeleton dynamics. This gene is highly expressed in the brain, but also in the heart and lungs to a lesser extent (51), thus partially overlapping with tissues where *FXN* is expressed (1). Two other genes, *PRKACG* and *TJP2*, are located immediately upstream the 5′ end and downstream the 3′ end of the *FXN* locus, respectively, and show no differences in expression in FRDA patient cells (50). Information on the *FXN* locus interaction in *trans* is not yet available and is beyond the scope of our study. However, it would be interesting to analyse the expanded *FXN* localization in FRDA-relevant tissues, such as cardiomyocytes and sensory neurons, and assess expression levels of potential spatial *FXN*-interacting genes.

Here, we have focused on the nuclear location of and transcription from the expanded *FXN* gene at the single-cell level. We find an increased retention of the expanded *FXN* gene at the NL. We show that a higher proportion of expanded *FXN* alleles are inactive when compared with normal alleles and we have demonstrated reduced transcription from the expanded alleles that are active. We provide a single-cell quantification showing that expanded GAA repeats impair expression predominantly at the initiation stage of transcription, but also block RNAPII elongation. It is the combination of these factors which together reduce the levels of frataxin protein to cause disease. Our work provides new mechanistic insights into the molecular causes of FRDA which may extend to other genetic diseases mediated by repeat expansions within regions of non-coding DNA.

Materials and Methods

Cell culture

Epstein Barr virus-transformed lymphoblastoid cell lines GM14519 (from an FRDA clinically unaffected carrier, alleles with 1285 + 50 and 9 GAA repeats), GM15851 (from an unaffected individual with normal range of GAA repeats) and GM16209 (from an FRDA clinically affected individual, both alleles with 800 GAA repeats) were obtained from the Human Genetic Cell Repository of the Coriell Institute (USA) and propagated in RPMI 1640 medium supplemented with 15% fetal bovine serum (FBS) and 2 mM L-glutamine. HEK FRT cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS, 2 mM L-glutamine, 100 U/ml penicillin/streptomycin and 100 µg/ml Zeocin (Life Technologies). *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* clonal cell lines were propagated in complete DMEM medium (see above) supplemented with 100 µg/ml Hygromycin B (Life Technologies). Actinomycin D (ActD) time course was performed by adding dimethyl sulphoxide (DMSO) or 8 µg/ml ActD to the media of *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* clonal cell lines before lysing cells every 4 h up to 24 h for RNA extraction. *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells were treated with DMSO, nicotinamide (10 mM) or compound 106 (10 µM) for 48 h before RNA extraction and Immuno-FISH.

Vector construction and stable cell line generation

pBAC-*FXN-GAA-MS2-Luc* was generated from the pBAC-*FXN-GAA-Luc* vector (25) which carries the whole 80 kb *FXN* locus with exons 1–5b of the *FXN* gene, an insertion of the luciferase gene in exon 5a and an ∼310 GAA repeat expansion in intron 1. A selection/counter-selection homologous recombination protocol based on RedET recombination (GeneBridges) was used. The recombination using the rpsL-neo and the pSC101-BAD plasmids was carried out in two steps in *E. coli* according to the manufacturer's instructions. In the first step, a PCR product containing the rpsL-neo cassette flanked by 55 bp homology arms to either side of exon 2 was used to insert the cassette. The replacement of the rpsL-neo cassette in the second step of recombination by 24 MS2 Binding Sites in exon 2 was performed as follows. A plasmid containing exon 2 with an engineered *Bam*HI restriction site and flanking homology arms (147 bp upstream and 172 bp downstream of exon 2) was purchased from GeneArt (Life Technologies). The 24 MS2 Binding Sites were excised by *Bam*HI/*Bgl*II digestion from pCR4-24XMS2SL-stable (52) (Addgene plasmid 31865) and inserted in exon 2 at the *Bam*HI site (Exon2-MBS). The rpsL-neo cassette was replaced with a PCR product containing the Exon2-MBS sequence flanked by homology arms, generating pBAC-*FXN-GAA-MS2-Luc*. To generate pBAC-*FXN-MS2-Luc*, the ∼310 GAA repeats in intron 1 in pBAC-*FXN-MS2-Luc* were replaced by the rpsL-neo cassette amplified with primers carrying 58 bp homology arms to sequences immediately upstream and downstream of GAA repeats. The rpsL-neo was then replaced with a PCR product containing six GAA repeats amplified from the genomic DNA of GM15851 using GAA-F and GAA-R primers (1). Successful construction was confirmed by PCR, restriction enzyme digestion followed by pulsed-field gel electrophoresis and sequencing. *Cre-loxP*-mediated retrofitting of pBAC-*FXN-MS2-Luc* and pBAC-*FXN-GAA-MS2-Luc* vectors to pH-FRT-Hy was performed as previously described (53). Stable clonal cell lines were obtained by transfecting HEK FRT cells with either pBAC-*FXN-MS2-Luc* or pBAC-*FXN-GAA-MS2-Luc* together with Flp recombinase-encoding plasmid pOG44 (Life Technologies), generating the lines *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc*, respectively. Stable selection was performed with 100 µg/ml hygromycin (Life Technologies). Stable clones were expanded and characterized as previously described (25).

Amplification of GAA repeats

Genomic DNA from *FXN-GAA-MS2-Luc* and *FXN-MS2-Luc* clonal cell lines was isolated using Illustra Tissue and Cells

GenomicPrep Mini Spin Kit (GE Healthcare), according to the manufacturer's instructions. PCR amplification of the GAA repeat sequence was carried out on 200 ng of genomic DNA using primers 147F and 602R (54) and Expand Long Template DNA Polymerase (Roche) as previously described (1,55). PCR products were visualized on a 1% agarose gel. The GAA repeat sequences were sequenced to check for interruptions.

PCR, qPCR and copy number assay

Vector integration and copy number determination were performed as previously described (25). The integration of pBAC-*FXN-MS2-Luc* and pBAC-*FXN-GAA-MS2-Luc* at the docking site was assessed by PCR analysis using primers pSV40-F and Hygro-R. Copy number was determined by real-time PCR, using the relative standard curve method. The upstream region of the GAA repeats in intron 1 of the *FXN* gene was amplified and normalized by *GAPDH*. The number of transgene copies was determined by comparison with the acceptor cell line HEK FRT as reference sample, which carries three endogenous *FXN* loci (25).

qRT-PCR

Total RNA from HEK FRT, *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* clonal cell lines and from the lymphoblastoid-derived cell lines GM15851, GM14519 and GM16209 was extracted using RNeasy Mini Kit (Qiagen) and treated with RNase-free DNase (Qiagen). cDNA was synthesized from 1 µg of total RNA using random primers (Life Technologies) and SuperScript III Reverse Transcriptase (Life Technologies) as previously described (25). Three independent cDNA samples per cell line were quantified in triplicate by real-time PCR using the SYBR Green PCR Master Mix (Applied Biosystems). *FXN-MS2-Luc* mRNA expression was determined using primers qFXN-Luc-F and qFXN-Luc-R (25) normalized by *GAPDH*. *FXN* mRNA expression was detected using primers FXN-F and FXN-R (8) normalized by β-actin.

Luciferase assay

To assay luciferase expression, 1.5×10^6 cells of *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* clonal cell lines were used as previously described (25). Cells were seeded in 6-well plates, washed in phosphate-buffered saline (PBS) and lysed in lysis buffer (25 mM Tris–PO_4, pH 7.8, 2 mM CDTA, 10% glycerol and 1% Triton-X 100) for 20 min at 4°C. Seventy-five microlitres of lysates were mixed with 100 µl of luciferase assay buffer (15 mM $MgSO_4$, 15 mM KPO_4, pH 7.8, 4 mM ethylene glycol tetraacetic acid, pH 7.8, 2 mM adenosine 5′-triphosphate and 2 mM dithiothreitol) and 50 µl of D-luciferin (0.3 mg/ml). The relative light units of luciferase of each cell line were determined using the Dynex MLX 96 Well Plate Luminometer and were normalized by total protein concentration, determined using bicinchoninic acid solution (BCA, Sigma).

Probes for DNA FISH

Human *FXN* gene was detected using either G248P87590H8 or G248P89980G8 fosmid clones obtained from BACPAC Resources Center (Children's Hospital, Oakland Research Institute). Probes were labelled by nick translation with digoxigenin (Roche). Briefly, 1 µg of DNA from each fosmid clone was treated with 200 ng RNase (Sigma) for 30 min at 37°C. After RNA removal, each probe was mixed with 5 µl of nick translation buffer [(0.5 M Tris–HCl, pH 8.0, 50 mM $MgCl_2$, 0.5 mg/ml bovine serum albumin (BSA)], 5 µl of β-mercaptoethanol, 5 µl of deoxynucleotide triphosphate mix (0.5 mM deoxyadenosine triphosphate, 0.5 mM deoxyguanosine triphosphate, 0.5 mM deoxycytidine triphosphate), 1 nmol of digoxigenin, 6 units of DNAse I (Roche) and 10 units of DNA polymerase I (Life Technologies) and incubated for 2 h at 16°C. Unincorporated nucleotides were removed using Illustra G-50 columns (GE Healthcare), according to the manufacturer's instructions. In carrier cells, the expanded allele was distinguished from the WT allele using biotinylated $(GAA)_{15}$ or $(TTC)_{15}$ oligos (Biomers). Chromosome 9 was detected using a green XCyting 9 whole chromosome paint for HSA9 (MetaSystems).

2D DNA FISH

DNA FISH was performed on GM14519 carrier cell line as described previously with slight modifications (56). Cells were fixed in 3:1 methanol–acetic acid and treated with 100 µg/ml RNase for 30 min at 37°C, washed in 2× SSC and dehydrated. Cells were denatured in 70% formamide/2× SSC at 72°C for 2 min and dehydrated. Labelled oligos (100 ng) were denatured in FISH hybridization buffer (Kreatech) at 90°C for 8 min and immediately allowed to cool on ice. Labelled fosmid clones (100 ng) were precipitated with 3 µg of human Cot-1 DNA (Life Technologies) and 20 µg of salmon sperm DNA (Sigma), denatured in FISH hybridization buffer at 90°C for 8 min and were preannealed at 37°C for 8 min. Whole chromosome 9 paint (4 µl) was denatured at 75°C for 2 min and combined with 8 µl of denatured oligo and fosmid and applied onto the slide. Slides were hybridized overnight at 37°C. After hybridization, slides were washed in 0.4× SSC/0.3% Igepal for 2 min and then 1 min at room temperature. Slides were blocked in 3% BSA/4× SSC for 30 min, biotin was detected with streptavidin Cy3 (Jackson ImmunoResearch Laboratories, 1/200 dilution) and digoxygenin was detected with sheep anti-digoxigenin FITC (Roche, 1/50 dilution) followed by rabbit anti-sheep FITC (Vector Laboratories, 1/100 dilution) or mouse anti-digoxigenin (Jackson ImmunoResearch Laboratories, 1/100 dilution) followed by Alexa Fluor 647 goat anti-mouse IgG (Life Technologies, A-21236, 1/200 dilution) when the green whole chromosome 9 paint was used. Cells were mounted in Vectashield (Vector Laboratories) with 1 µg/ml 4′,6-diamidino-2-phenylindole (DAPI) counterstain.

3D DNA FISH

3D DNA FISH was performed on GM14519 carrier cell line as described previously with slight modifications (56). Cells were washed in PBS and allowed to settle on poly-L-lysine-treated coverslips for 5 min. Cells were fixed in 4% paraformaldehyde (PFA)/25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) for 15 min and permeabilized in 0.5% Triton X-100 (Calbiochem) in PBS for 10 min. RNA was removed with 100 µg/ml RNase in 2× SSC for 1 h at 37°C. Cells were denatured in 3 N HCl for 20 min and neutralized in ice-cold PBS. Probes were prepared as in the previous section, and slides were hybridized overnight at 37°C. Cells were washed twice in 2× SSC at 37°C, once in 1× SSC at RT for 30 min, and blocked in 3% BSA/4× SSC for 1 h. Probes were detected as in the previous section. Cells were washed between layers in 4× SSC with 0.05% Tween-20. Coverslips were mounted in Vectashield with 1 µg/ml DAPI counterstain.

Immuno-FISH

Immuno-FISH was performed on *FXN-MS2-Luc* and *FXN-GAA-MS2-Luc* cells and on the lymphoblastoid cell lines GM14519,

GM15851, and GM16209 as described previously with slight modifications (56). Adherent cells were seeded on poly-L-lysine-treated coverslips and allowed to grow overnight. Suspension cells were washed in PBS and allowed to settle on poly-L-lysine-treated coverslips for 5 min. Cells were fixed in 4% PFA/25 mM HEPES for 15 min and permeabilized in 0.5% Triton X-100 for 10 min. Non-specific sites were blocked with 10% FBS. Antibodies were prepared in blocking solution. The NL was detected with goat anti-lamin B1 (Santa Cruz, sc-6217, 1/200 dilution). Nuclear speckles were detected with mouse anti-SC35 (Sigma, S4045, 1/1000 dilution). The nucleolus was detected with mouse anti-nucleolin (Abcam, ab13541, 1/25 dilution). Secondary antibodies used were Alexa Fluor 647 goat anti-mouse IgG (Life Technologies, A-21236, 1/1000 dilution) and Alexa Fluor 647 rabbit anti-goat IgG (Life Technologies, A-21446, 1/500 dilution). After protein detection, the slides were post-fixed in 4% PFA/25 mM HEPES at RT for 10 min. Cells were treated with 100 µg/ml RNase for 1 h at 37°C, denatured in 3 N HCl for 20 min and neutralized in ice-cold PBS. Probes were prepared as in the 2D DNA FISH section, and slides were hybridized overnight at 37°C. Cells were washed twice in 2× SSC at 37°C, once in 1× SSC at RT and blocked in 3% BSA/4× SSC. Probes were detected as in the 3D DNA-FISH section.

RNA FISH

RNA FISH was performed on HEK FRT and *FXN-MS2-Luc* cell lines as previously described (57) with minor modifications. Briefly, cells were allowed to grow on poly-L-lysine-treated coverslips overnight, washed in PBS and fixed in 4% formaldehyde (Electron Microscopy Science) for 10 min at RT. Cells were permeabilized in 0.5% Triton X-100 for 6 min and washed in 50% formamide/2× SSC for 5 min. Cells were hybridized overnight at 37°C with 10 ng Cy3-MS2 DNA probe (58) in 40 µl hybridization mix containing 2× SSC, 3% BSA, 10% dextran sulphate (Sigma), 2 mM vanadyl-ribonucleoside complex (New England Biolabs), 40 µg tRNA (Roche) and 50% formamide. Cells were washed once in 50% formamide/2× SSC at 37°C for 10 min, once in 2× SSC for 10 min at 37°C and once in 2× SSC for 5 min at RT. Cells were washed briefly in PBS and mounted in Vectashield with 1 µg/ml DAPI counterstain.

RNA FISH on GM15851 and GM16209 cell lines was performed as described previously (59) with minor modifications. Custom Stellaris FISH probes TSS (20 oligos), UpGAA (41 oligos) and DownGAA (41 oligos) were designed to hybridize with the regions TSS-exon1, upstream and downstream of the GAA repeats, respectively (Supplementary Material, Table S1), according to the manufacturer's instructions. Cells were washed in PBS and allowed to settle on poly-L-lysine-treated coverslips for 5 min. Cells were fixed in 3.7% formaldehyde/1× PBS for 10 min, permeabilized in 70% ethanol at 4°C for 1 h and washed in 10% formamide/2× SSC at RT for 5 min. Hybridization was performed overnight at 42°C with probes TSS (50 nM), UpGAA (50 nM) and DownGAA (25 nM) in 100 µl hybridization buffer containing 100 mg/ml dextran sulphate, 10% formamide in 2× SSC. Cells were washed in 10% formamide/2× SSC at 42°C for 10 min. Cells were washed briefly in PBS and mounted in Vectashield with 1 µg/ml DAPI counterstain.

Fluorescence recovery after photobleaching

FRAP experiments were performed using the UltraView Spinning Disk system (PerkinElmer) mounted on an IX81 microscope (Olympus) equipped with a 100× 1.3 NA objective, a FRAP unit and an electron multiplying charge-coupled device (EMCCD) camera (Hamamatsu Photonics) and driven by Volocity software (PerkinElmer). HEK FRT and *FXN-MS2-Luc* cell lines were seeded on collagen-coated glass-bottom dishes (MatTek) in complete DMEM medium (see above) and transiently transfected with NES-YFP-MS2-NLS (60) (a kind gift from Y. Shav-Tal). For nuclear detection, NucBlue Live reagent was added to the cells 20 min before imaging. Cells were maintained at 37°C in a 5% CO_2 humidified atmosphere using an on-scope incubator and an objective heater (Tokai). Fluorescent spots inside the nucleus indicating active *FXN-MS2-Luc* transcription sites were detected with the YFP channel (514 nm laser). The nucleus was detected with the DAPI channel (405 nm laser). z-stacks were acquired at 0.25 µm intervals to cover the entire nuclear volume of each cell. Before bleaching, three z-stacks of each cell were taken. The active transcription site was bleached for 200 ms with the 514 nm laser power set to 100%. Post-bleach recovery z-stacks images were acquired with an exposure time of 50 ms every 20 s for 20 min.

Imaging acquisition and analysis

Images of fixed cells were obtained using the DV Elite system based on an Olympus IX71 fully motorized wide-field deconvolution inverted microscope with a 100× objective 1.40 numerical aperture (NA) fitted with a CoolSNAP HQ^2 cooled charge-coupled device (CCD) camera (Photometrics) driven by SoftWoRx 5.0 software (Applied Precision). Several cell positions were chosen randomly and recorded using the motorized stage. z-stacks were acquired at 0.2 µm intervals to cover the entire volume of each cell. Distance measurements between FISH signals and the NP were calculated with NEMO software (61). Briefly, a region of interest (ROI) was defined around each nucleus using an image stack from the DAPI channel. The software applied several segmentation filters (Median 3D, TopHat 3D, mathematical morphology 3D) on raw image stacks for object detection. Objects were detected by computing seeds as the brightest maximum local pixels above a global threshold, and a local threshold was then computed to aggregate neighbouring pixels to the seed. For each cell, the distance of each allele to the NP was normalized by the local ray between the nuclear centre and periphery passing through the FISH signal. For relative positional analysis, image stacks were deconvolved using the SoftWoRx 5.0 software. Association of FISH signals with subnuclear domains or chromosome territories was analysed using ImageJ (http://imagej.nih.gov/ij/): image stacks were analysed manually, and association was determined by eye when *FXN* FISH signals were touching or superimposing the signal from a specific subnuclear domain at a given z-section. Images are maximum projections of z-stacks covering the signals. Contrast-stretch and gamma adjustments were made using Adobe Photoshop CS5 only for display in Figures 3B, F and 4A.

RNA MS2 FISH signals were detected as described previously (29) using StarSearch software (http://rajlab.seas.upenn.edu/StarSearch/launch.html). Briefly, the program applied Laplacian of Gaussian filters in three dimensions, taking into consideration the size of the fluorescent spots to remove the non-uniform background while enhancing particles. After filtering, the program calculated the number of spots detected for all possible thresholds. After plotting this data, the graph showed a plateau region corresponding to a broad range of thresholds over which the spot count did not vary significantly. Spot counts calculated with these thresholds matched the number of spots identified by eye. In HEK FRT cells and cells from the *FXN-MS2-Luc* cell lines that did not express the transgene, the plateau region was not observed. The DIC channel was used to determine the cell edges.

In RNA FISH experiments on GM15851 and GM16209 cell lines, several cell positions were chosen randomly. z-stacks were acquired at 0.2 µm intervals to cover the entire volume of each nucleus. Images were analysed using ImageJ. For the analysis of nascent RNA FISH signals using the UpGAA probe only, a threshold was defined for both cell lines to eliminate non-specific background. RNA fluorescent spots were identified either at the nuclear rim or interior using the DAPI channel and their intensity measured using the 3D Object Counter plugin (62). Identification of RNA FISH signals using probes TSS, UpGAA and DownGAA was performed as described above. In addition to threshold determination, to distinguish true signals from background, the downstream signals were scored when co-localized with the upstream signals. Because the TSS probe hybridizes to exon 1, it could potentially label mature *FXN* mRNA. However, due to its short length spanning, the TSS probe did not generate more than two RNA FISH signals per cell. Therefore, all TSS RNA FISH signals were identified as nascent transcript accumulating at the transcription sites.

For live-cell experiments, HEK FRT and *FXN-MS2-Luc* cell lines were seeded on collagen-coated glass-bottom dishes (MatTek) in complete DMEM medium (see above) and transiently transfected with NES-YFP-MS2-NLS (60). For live-cell imaging of transcription sites, cells were imaged in four dimensions (three dimensions over time) using the DV Elite system based on an Olympus IX71 microscope with a 100× objective 1.40 NA fitted with an Evolve EMCCD camera (Photometrics) driven by SoftWoRx 5.0 software under physiological conditions inside an incubator at 37°C and 5% CO_2.

FRAP image sequences were analysed with Volocity software (PerkinElmer). Tracking of fluorescence spots was performed using the manual tracking function on Volocity at each time point. Relative intensity of each time point was calculated using

$$I_{\mathrm{rel}} = (T_0 - B_0) \times \frac{I_t - B_t}{T_t - B_t} \times (I_0 - B_0)$$

where T_0 and T_t are the average intensity of an arbitrary area in the nucleus before and after bleaching, respectively, and I_0 and I_t are the average intensity of the ROI before and after bleaching, respectively. The background was subtracted from all measured values. The formula accounted for fluorescence loss due to bleaching monitoring, and finally fluorescence intensity was normalized to 1 (63). Data from five cells per lines were collected in three independent experiments, and the averaged FRAP measurements were fitted with exponential curves. The diffusion of MS2-YFP protein during FRAP analysis was ignored as the diffusion rate of free MS2-YFP is very rapid, whereas the bound MS2-YFP is associated with high affinity to the mRNA and does not detach, diffuse and bind again at the transcription site (30).

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

We thank Y. Shav-Tal for the gift of NES-YFP-MS2-NLS construct, J.R. Chubb, R. Parton and T. Weil for helpful discussions.

Conflict of Interest statement. None declared.

Funding

This work was supported by Ataxia UK (fellowship no. 7125 to M. M.L.), the Friedreich's Ataxia Research Alliance (FARA) (to M.M.L. and R.W.-M.), BabelFAmily (to M.M.L.), Associazione Italiana per la lotta alle Sindromi Atassiche (AISA) (to M.M.L.), the European Union 7th Framework Program EFACTS (grant agreement no. 242193) (to R.W.-M.). M.M.L. is an Ataxia UK research fellow co-funded by FARA and AISA. A.M.S. was supported by Fundação para a Ciência e a Tecnologia PhD studentship (SFRH/BD/61048/2009). Advanced microscopy at Micron Oxford was supported by a Wellcome Trust Strategic Award (091911).

References

1. Campuzano, V., Montermini, L., Moltò, M.D., Pianese, L., Cossée, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A. *et al.* (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science*, **271**, 1423–1427.
2. Cossée, M., Puccio, H., Gansmuller, A., Koutnikova, H., Dierich, A., LeMeur, M., Fischbeck, K., Dollé, P. and Kœnig, M. (2000) Inactivation of the Friedreich ataxia mouse gene leads to early embryonic lethality without iron accumulation. *Hum. Mol. Genet.*, **9**, 1219–1226.
3. Sakamoto, N., Chastain, P.D., Parniewski, P., Ohshima, K., Pandolfo, M., Griffith, J.D. and Wells, R.D. (1999) Sticky DNA: self-association properties of long GAA·TTC repeats in R·R·Y triplex structures from Friedreich's ataxia. *Mol. Cell*, **3**, 465–475.
4. Bidichandani, S.I., Ashizawa, T. and Patel, P.I. (2014) The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. *Am. J. Hum. Genet.*, **62**, 111–121.
5. Grabczyk, E., Mancuso, M. and Sammarco, M.C. (2007) A persistent RNA·DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP in vitro. *Nucleic Acids Res.*, **35**, 5351–5359.
6. Groh, M., Lufino, M.M.P., Wade-Martins, R. and Gromak, N. (2014) R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and Fragile X Syndrome. *PLoS Genet.*, **10**, e1004318.
7. Saveliev, A., Everett, C., Sharpe, T., Webster, Z. and Festenstein, R. (2003) DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing. *Nature*, **422**, 909–913.
8. Herman, D., Jenssen, K., Burnett, R., Soragni, E., Perlman, S.L. and Gottesfeld, J.M. (2006) Histone deacetylase inhibitors reverse gene silencing in Friedreich's ataxia. *Nat. Chem. Biol.*, **2**, 551–558.
9. Greene, E., Mahishi, L., Entezam, A., Kumari, D. and Usdin, K. (2007) Repeat-induced epigenetic changes in intron 1 of the frataxin gene and its consequences in Friedreich ataxia. *Nucleic Acids Res.*, **35**, 3383–3390.
10. Al-Mahdawi, S., Pinto, R.M., Ismail, O., Varshney, D., Lymperi, S., Sandi, C., Trabzuni, D. and Pook, M. (2008) The Friedreich ataxia GAA repeat expansion mutation induces comparable epigenetic changes in human and transgenic mouse brain and heart tissues. *Hum. Mol. Genet.*, **17**, 735–746.
11. Evans-Galea, M.V., Carrodus, N., Rowley, S.M., Corben, L.A., Tai, G., Saffery, R., Galati, J.C., Wong, N.C., Craig, J.M., Lynch, D.R. *et al.* (2012) FXN methylation predicts expression and clinical outcome in Friedreich ataxia. *Ann. Neurol.*, **71**, 487–497.

12. Punga, T. and Bühler, M. (2010) Long intronic GAA repeats causing Friedreich ataxia impede transcription elongation. *EMBO Mol. Med.*, **2**, 120–129.
13. De Biase, I., Chutake, Y.K., Rindler, P.M. and Bidichandani, S.I. (2009) Epigenetic silencing in Friedreich ataxia is associated with depletion of CTCF (CCCTC-binding factor) and antisense transcription. *PLoS ONE*, **4**, e7914.
14. Kumari, D., Biacsi, R.E. and Usdin, K. (2011) Repeat expansion affects both transcription initiation and elongation in Friedreich ataxia cells. *J. Biol. Chem.*, **286**, 4209–4215.
15. Chutake, Y.K., Costello, W.N., Lam, C. and Bidichandani, S.I. (2014) Altered nucleosome positioning at the transcription start site and deficient transcriptional initiation in Friedreich ataxia. *J. Biol. Chem.*, **289**, 15194–15202.
16. Dillon, N. (2008) The impact of gene location in the nucleus on transcriptional regulation. *Dev. Cell*, **15**, 182–186.
17. Takizawa, T., Meaburn, K.J. and Misteli, T. (2008) The meaning of gene positioning. *Cell*, **135**, 9–13.
18. Ferrai, C., de Castro, I.J., Lavitas, L., Chotalia, M. and Pombo, A. (2010) Gene positioning. *Cold Spring Harb. Perspect. Biol.*, **2**, a000588.
19. Therizols, P., Illingworth, R.S., Courilleau, C., Boyle, S., Wood, A.J. and Bickmore, W.A. (2014) Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science*, **346**, 1238–1242.
20. Finlan, L.E., Sproul, D., Thomson, I., Boyle, S., Kerr, E., Perry, P., Ylstra, B., Chubb, J.R. and Bickmore, W.A. (2008) Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet.*, **4**, e1000039.
21. Reddy, K.L., Zullo, J.M., Bertolino, E. and Singh, H. (2008) Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature*, **452**, 243–247.
22. Zullo, J.M., Demarco, I.A., Piqué-Regi, R., Gaffney, D.J., Epstein, C.B., Spooner, C.J., Luperchio, T.R., Bernstein, B.E., Pritchard, J.K., Reddy, K.L. *et al.* (2014) DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina. *Cell*, **149**, 1474–1487.
23. Kumaran, R.I. and Spector, D.L. (2008) A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence. *J. Cell Biol.*, **180**, 51–65.
24. Kind, J. and van Steensel, B. (2010) Genome–nuclear lamina interactions and gene regulation. *Curr. Opin. Cell Biol.*, **22**, 320–325.
25. Lufino, M.M.P., Silva, A.M., Németh, A.H., Alegre-Abarrategui, J., Russell, A.J. and Wade-Martins, R. (2013) A GAA repeat expansion reporter model of Friedreich's ataxia recapitulates the genomic context and allows rapid screening of therapeutic compounds. *Hum. Mol. Genet.*, **22**, 5173–5187.
26. Xu, C., Soragni, E., Chou, C.J., Herman, D., Plasterer, H.L., Rusche, J.R. and Gottesfeld, J.M. (2009) Chemical probes identify a role for histone deacetylase 3 in Friedreich's ataxia gene silencing. *Chem. Biol.*, **16**, 980–989.
27. Chan, P.K., Torres, R., Yandim, C., Law, P.P., Khadayate, S., Mauri, M., Grosan, C., Chapman-Rothe, N., Giunti, P., Pook, M. *et al.* (2013) Heterochromatinization induced by GAA-repeat hyperexpansion in Friedreich's ataxia can be reduced upon HDAC inhibition by vitamin B3. *Hum. Mol. Genet.*, **22**, 2662–2675.
28. Fusco, D., Accornero, N., Lavoie, B., Shenoy, S.M., Blanchard, J.M., Singer, R.H. and Bertrand, E. (2003) Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. *Curr. Biol.*, **13**, 161–167.
29. Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A. and Tyagi, S. (2008) Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods*, **5**, 877–879.
30. Yunger, S., Rosenfeld, L., Garini, Y. and Shav-Tal, Y. (2010) Single-allele analysis of transcription kinetics in living mammalian cells. *Nat. Methods*, **7**, 631–633.
31. Lockhart, D.J. and Winzeler, E.A. (2000) Genomics, gene expression and DNA arrays. *Nature*, **405**, 827–836.
32. Akhtar, W., De Jong, J., Pindyurin, A.V., Pagie, L., Meuleman, W., De Ridder, J., Berns, A., Wessels, L.F.A., Van Lohuizen, M. and van Steensel, B. (2013) Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell*, **154**, 914–927.
33. Meaburn, K.J. and Misteli, T. (2008) Locus-specific and activity-independent gene repositioning during early tumorigenesis. *J. Cell Biol.*, **180**, 39–50.
34. Takizawa, T., Gudla, P.R., Guo, L., Lockett, S. and Misteli, T. (2008) Allele-specific nuclear positioning of the monoallelically expressed astrocyte marker GFAP. *Genes Dev.*, **22**, 489–498.
35. Shopland, L.S., Johnson, C.V., Byron, M., McNeil, J. and Lawrence, J.B. (2003) Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains: evidence for local euchromatic neighborhoods. *J. Cell Biol.*, **162**, 981–990.
36. Chambeyron, S. and Bickmore, W.A. (2004) Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev.*, **18**, 1119–1130.
37. Brown, J.M., Leach, J., Reittie, J.E., Atzberger, A., Lee-Prudhoe, J., Wood, W.G., Higgs, D.R., Iborra, F.J. and Buckle, V.J. (2006) Coregulated human globin genes are frequently in spatial proximity when active. *J. Cell Biol.*, **172**, 177–187.
38. Zhang, L.-F., Huynh, K.D. and Lee, J.T. (2007) Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. *Cell*, **129**, 693–706.
39. Zhao, R., Bodnar, M.S. and Spector, D.L. (2009) Nuclear neighborhoods and gene expression. *Curr. Opin. Genet. Dev.*, **19**, 172–179.
40. Levsky, J.M. and Singer, R.H. (2003) Gene expression and the myth of the average cell. *Trends Cell Biol.*, **13**, 4–6.
41. Kim, E., Napierala, M. and Dent, S.Y.R. (2011) Hyperexpansion of GAA repeats affects post-initiation steps of FXN transcription in Friedreich's ataxia. *Nucleic Acids Res.*, **39**, 8366–8377.
42. Colak, D., Zaninovic, N., Cohen, M.S., Rosenwaks, Z., Yang, W.-Y., Gerhardt, J., Disney, M.D. and Jaffrey, S.R. (2014) Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. *Science*, **343**, 1002–1005.
43. Schooley, A., Vollmer, B. and Antonin, W. (2012) Building a nuclear envelope at the end of mitosis: coordinating membrane reorganization, nuclear pore complex assembly, and chromatin de-condensation. *Chromosoma*, **121**, 539–554.
44. Kind, J., Pagie, L., Ortabozkoyun, H., Boyle, S., de Vries, S.S., Janssen, H., Amendola, M., Nolen, L.D., Bickmore, W.A. and van Steensel, B. (2013) Single-cell dynamics of genome-nuclear lamina interactions. *Cell*, **153**, 178–192.
45. Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W. *et al.* (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, **453**, 948–951.
46. Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W.M., Solovei, I., Brugman, W., Gräf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M. *et al.* (2010) Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell*, **38**, 603–613.

47. Chubb, J.R., Boyle, S., Perry, P. and Bickmore, W.A. (2002) Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr. Biol.*, **12**, 439–445.
48. Klesert, T., Otten, A., Bird, T. and Tapscott, S. (1997) Trinucleotide repeat expansion at the myotonic dystrophy locus reduces expression of DMAHP. *Nat. Genet.*, **16**, 402–406.
49. Thornton, C., Wymer, J., Simmons, Z., McClain, C. and Moxley, R. (1997) Expansion of the myotonic dystrophy CTG repeat reduces expression of the flanking DMAHP gene. *Nat. Genet.*, **16**, 407–409.
50. Bayot, A., Reichman, S., Lebon, S., Csaba, Z., Aubry, L., Sterkers, G., Husson, I., Rak, M. and Rustin, P. (2013) *cis*-silencing of PIP5K1B evidenced in Friedreich's ataxia patient cells results in cytoskeleton anomalies. *Hum. Mol. Genet.*, **22**, 2894–2904.
51. Volpicelli-Daley, L.A., Lucast, L., Gong, L.-W., Liu, L., Sasaki, J., Sasaki, T., Abrams, C.S., Kanaho, Y. and De Camilli, P. (2010) Phosphatidylinositol-4-phosphate 5-kinases and phosphatidylinositol 4,5-bisphosphate synthesis in the brain. *J. Biol. Chem.*, **285**, 28708–28714.
52. Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S.M., Singer, R.H. and Long, R.M. (1998) Localization of ASH1 mRNA particles in living yeast. *Mol. Cell*, **2**, 437–445.
53. Lufino, M.M.P., Manservigi, R. and Wade-Martins, R. (2007) An S/MAR-based infectious episomal genomic DNA expression vector provides long-term regulated functional complementation of LDLR deficiency. *Nucleic Acids Res.*, **35**, e98.
54. Montermini, L., Andermann, E., Labuda, M., Richter, A., Pandolfo, M., Cavalcanti, F., Pianese, L., Iodice, L., Farina, G., Monticelli, A. *et al.* (1997) The Friedreich ataxia GAA triplet repeat: premutation and normal alleles. *Hum. Mol. Genet.*, **6**, 1261–1266.
55. Al-Mahdawi, S., Pinto, R.M., Ruddle, P., Carroll, C., Webster, Z. and Pook, M. (2004) GAA repeat instability in Friedreich ataxia YAC transgenic mice. *Genomics*, **84**, 301–310.
56. Brown, J.M., Green, J., das Neves, R.P., Wallace, H.A.C., Smith, A.J.H., Hughes, J., Gray, N., Taylor, S., Wood, W.G., Higgs, D.R. *et al.* (2008) Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *J. Cell Biol.*, **182**, 1083–1097.
57. Chartrand, P., Bertrand, E., Singer, R.H. and Long, R.M. (2000) Long, sensitive and high-resolution detection of RNA in situ. *Methods Enzymol.*, **318**, 493–506.
58. Hocine, S., Raymond, P., Zenklusen, D., Chao, J.A. and Singer, R. H. (2013) Single-molecule analysis of gene expression using two-color RNA labeling in live yeast. *Nat. Methods*, **10**, 119–121.
59. Raj, A. and Tyagi, S. (2010) Detection of individual endogenous RNA transcripts in situ using multiple singly labeled probes. In Nils, G.M. (ed), *Methods in Enzymology*. Academic Press, New York, USA, Vol. 472, pp. 365–382.
60. Ben-Ari, Y., Brody, Y., Kinor, N., Mor, A., Tsukamoto, T., Spector, D.L., Singer, R.H. and Shav-Tal, Y. (2010) The life of an mRNA in space and time. *J. Cell Sci.*, **123**, 1761–1774.
61. Iannuccelli, E., Mompart, F., Gellin, J., Lahbib-Mansais, Y., Yerle, M. and Boudier, T. (2010) NEMO: a tool for analyzing gene and chromosome territory distributions from 3D-FISH experiments. *Bioinformatics*, **26**, 696–697.
62. Bolte, S. and Cordelières, F.P. (2006) A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.*, **224**, 213–232.
63. Dundr, M. and Misteli, T. (2003) Measuring dynamics of nuclear proteins by photobleaching. *Curr. Protoc. Cell Biol.*, **18**, 13.5.1–13.5.18.